

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 14:58:11 ; Search time 90.5 Seconds
(without alignments)
12118.688 Million cell updates/sec

Title: US-09-323-597C-1

Perfect score: 3179

Sequence: 1 99cggaggcggcggagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p,model -DEV=xlp
-Q=/cgn2/1/USPTO_spool_p/US09333597/runat_06072004_111728_29658/app_query.fasta_1.1927
-DB=SPREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US093233597 @CNG 1 1 99 @runat_06072004_111728_29658 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEFTIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_nhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp Vertebrate: *
14: sp Unclassified: *
15: sp_rvirus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2711	85.3	492 4	Q96t73 homo sapien

ID	Q96T73	PRELIMINARY;	PRT;	492 AA.
AC	Q96T73;			
DT	01-DEC-2001 (T-EMBLrel. 19, Created)			
DT	01-DEC-2001 (T-EMBLrel. 19, Last sequence update)			
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)			
DE	Epitheliasin.			
GN	TMPSRS2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21223025; PubMed=11322890;			
RA	Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,			
RA	Hoidal J.R.;			
RT	"Cloning and characterization of the cDNA and gene for human			
RT	epitheliasin."			
RL	Eur. J. Biochem. 268:2687-2699(2001).			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
DR	EMBL; AF329454; AAK5359.1; -			
DR	RSSP; P09761; IAN1.			
DR	GO; GO:0016020; C:membrane; IEA.			

ALIGNMENTS

RESULT 1

ID	Q96T73	PRELIMINARY;	PRT;	492 AA.
AC	Q96T73;			
DT	01-DEC-2001 (T-EMBLrel. 19, Created)			
DT	01-DEC-2001 (T-EMBLrel. 19, Last sequence update)			
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)			
DE	Epitheliasin.			
GN	TMPSRS2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21223025; PubMed=11322890;			
RA	Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,			
RA	Hoidal J.R.;			
RT	"Cloning and characterization of the cDNA and gene for human			
RT	epitheliasin."			
RL	Eur. J. Biochem. 268:2687-2699(2001).			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
DR	EMBL; AF329454; AAK5359.1; -			
DR	RSSP; P09761; IAN1.			
DR	GO; GO:0016020; C:membrane; IEA.			

DR GO: 0004263; F:chymotrypsin activity; IEA.
DR GO: 0008233; F:peptidase activity; IEA.
DR GO: 0005044; F:scavenger receptor activity; IEA.
DR GO: 0004295; F:trypsin activity; IEA.
DR GO: 0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR002172; LDL receptor A.
DR InterPro: IPR001254; Peptidase S1.
DR InterPro: IPR001344; Peptidase S1A.
DR InterPro: IPR001190; Srcr receptor.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLA; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00020; Tryp SPC; 1.
DR PROSITE: PS01209; LDLA_1; 1.
DR PROSITE: PS00068; LDLA_2; 1.
DR PROSITE: PS0287; SRCR_2; 1.
DR PROSITE: PS0240; TRYPSIN DOM; 1.
DR PROSITE: PS00134; TRYPSIN HIS; 1.
DR PROSITE: PS00135; TRYPSIN SER; 1.
KW Hydroxylase; Protease; Serine protease.
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Alignment Scores:
Pred. No.: 2,76e-230 Length: 492
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 85.28% Indels: 0
DB: 4 Gaps: 0

US-09-323-597c-1 (1-1738) x Q96T73 (1-492)
QY 112 ATGGCTTTGAACCTCAGCGTCACACAGCTATTGGACCTTACTATGAAACACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGACAGCCCACTGTGGTCCCACTGTCTACAGGPGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CGGGCTCAGTACTACCGCTCCCGCGGCCAGTACGCCCGCGGGCTCTCGAGCGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCACCCCGCTCTGTCACAGCGCCCAATCCCATCCGGGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTACAGACCACTGTGCATCACCCTGACCTGGGACCTTCTCTGGGAGCTCGCTG 411
DB 81 ThrLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCTGGCTCTCTGGAAGTTTCATGGGACAGTGTCTCAACTCTGGGATAGATGC 471
DB 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACCTGCAATCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGTGT 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGACGAGAACTCGTGTGTCTCGCTCTACGGACCAACTTCTATCTCTCAGGTG 591
DB 141 GlyGlyLysGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGGAAGCTCTGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GGGCGGGCGCTGACGGGACATGGCTATAGATAATTTTACTCTAGTCCCAAGGATA 711
DB 181 GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPherySerSerGlnGlyIle 200

QY 712 GTGGATGACAGGGATCCACAGCTTTATGAACTGAACACAGTGCCTGGCAATTCGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTACACAGTGTATCCCTCTTCAAAAGCACTGTTCTTCTTTCAGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCGCCAGAGAGAGATTGTGGCGCGCAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGCTGGCCCTGGCAGGTTCAGCTGCAGCTGCAGAACTGCACAGC 951
DB 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCCGAGGCTCCATCATCATCCCCCGAGTGTGTCAGCGCGCCCATCGCTGTGAAAAA 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGCGCATTTGGAGCGCATTTGGGGATTTTGAGCAATCTTTCATG 1071
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGGATACCAAGTAGAAAAAGTGTCTCTCATCCAAATTTATGACTCCCAAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGACATGACATTCGCTGATGAGAGCTGCGAAGCTCTGACTTTCAACGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACAGTGTCTGCCCAACCCAGCGCATGTGTCAGCGCCAGCAACAGCTCTGCTGG 1251
DB 361 ValLysProValCysLeuLeuProAsnProGlyMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGCCACCGAGAGAGAAAGGAGAGCTCAGAAAGTCTGAACCTGCC 1311
DB 381 IleSerGlyTyrGlyValThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTTAGACACAGAGATGCAACAGCAGATGTCTATGACAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCCATGATCTGTGCGGCTCTCTGAGGGAAACGTGATCTTTGCCAGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCGTCACTTCGAAGAACATATCTGGTGGCTGATAGGGGATACAAAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGCTGTGCCAAAGCTTACAGACAGAGTGTACGGGAATGTGATGTATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGATTATCGCAAAATGAGGCGCAGACGCG 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2
Q7TN04
ID Q7TN04 PRELIMINARY; PRT; 490 AA.
AC Q7TN04;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trpms62 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE TMPRSS2.
 GN TMPRSS2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tsuzuki S.;
 RT "TMPRSS2, Oct. '01";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBS databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB073550; BAB70683.1; --
 DR HSSP; P00761; IAN1.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR002172; LDL receptor A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Strc_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00068; LDLRA_2; 1.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;
 Alignment Scores:
 Pred. No.: 7,23e-175 Length: 490
 Score: 2083.00 Matches: 378
 Percent Similarity: 86.56% Conservative: 47
 Best Local Similarity: 76.99% Mismatches: 64
 Query Match: 65.52% Indels: 2
 DB: 11 Gaps: 2
 US-09-323-597c-1 (1-1738) x Q920K3 (1-490)
 QY 112 ATGGCTTTGAACCTACAGGTCCACACCACTATTGACCTTACTATGAAACCATGGATAC 171
 DB 1 MetAlaLeuAsnSerGlySerProGlyIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACGGAAACCCCTATCCCGCAGCCACGCTGTGTCCTCCACTGCTACAGGTGAT 231
 DB 21 GlnSerGluHisValTyrSerProArgProProValSerProSerGlyTyrAsnLeuTyr 40
 QY 232 CCGGCTCAGTACTACCCGTCCTCCCTGCCCCAGTACGCCCGGAGGCTCTCAGCAGGCT 291
 DB 41 ProAlaGlnSerCysProSerProValProGlnTyrAlaProArgValThrGlnAla 60
 QY 292 TCCACCCCGTGTCTGACGAGCCCAATCCCATCCCGGAGCAGGTGACCTCAAG 351
 DB 61 SerThrProAlaIleHisIleGlnProArgSer---SerGlyThrLeuCysThrSerTys 79
 QY 352 ACTAAGAAAGCAGCTGTGCATCACCTTGACCTCGGGACCTTCCCTCGTGGAGGTGGCTG 411
 DB 80 SerIlyslsSerMetLeuValAlaLeuAlaLeuGlyLeuSerProValAlaAlaVal 99
 QY 412 GCCTCGGCTTACTCTGGAAGTTTCATGGGCGAGCAAGTCTCCAACTCTGGGATAGATGC 471
 DB

Db 100 AlaAlaGlyLeuLeuTyrLysPheTrpAspSerLysCysSerSerSerGluMetGluCys 119
 QY 472 GACTCTCAGGTACTCTGATCAACCCCTCTAACTAGGTGTGATGCGTGTCACTACGCCCC 531
 DB 120 GlySerSerGlyThrCysIleSerSerSerLeuTyrCysAspGlyValAlaGlnCysPro 139
 QY 532 GCGCGGAGGACGAGAAATCGGTGTTCGGCTCTTACGGACCAAACTTCACTCTTCAGGTG 591
 DB 140 AsnGlyLysAspGluAsnArgCysValArgLeuTyrGlyThrSerPheThrLeuGlnVal 159
 QY 592 TACTCATCTCAGAGAGAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 651
 DB 160 TyrSerSerGlnArgLysAlaIleTyrProValCysGlnAspAspTrpAsnGluSerTyr 179
 QY 652 GCGCGGCGGCTCTCAGGGACATGGGTATAAGATAATTTTACTCTAGCAAGGAATA 711
 DB 180 GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnSerPheTyrSerSerGlnGlyLe 199
 QY 712 GTGATGACAGCGGATCCACAGCTTATGAACCTGAACACAGTGGCGGCAATGTGAT 771
 DB 200 ProAspGlnSerGlyAlaThrSerPheMetLysLeuAsnValSerAlaGlyAsnValAsp 219
 QY 772 ATCTATAAAACCTGTACCAAGTATGCCCTGTCTTCAAAGCAGTGGTTCCTTACGC 831
 DB 220 LeuTyrLysLysLeuTyrHisSerAspSerCysSerSerArgMetValValSerLeuArg 239
 QY 832 TGTATAGCTCGCGGTCAACTTGAACCTCAAGCGCCGAGCAGGATTTGGCGGCGGAG 891
 DB 240 CysIleGluCysGlyValArg---SerValArgArgGlnSerArgIleValGlyGlySer 258
 QY 892 AGCGCGCTCCCGGGGCTCGCCCTGGCAGGTGAGCCTGACGCTCCAGAACGTCCAGCTG 951
 DB 259 ThrAlaSerProGlyAspTrpProTyrGlnValSerLeuHisValGlnGlyIleHisVal 278
 QY 952 TGGGAGGCTTCATCATCACCCTCGAGTGTGACAGCCGCCCTCGCTGGTGGAAAAA 1011
 DB 279 CysGlyGlySerIleThrProGluTyrIleValThrAlaArgHisCysValGluGlu 298
 QY 1012 CCTCTTAACAAATCCATGCGCATTCGACGGCATTTGCGGGATTTTGACACAACTTCTTCATG 1071
 DB 299 ProLeuSerSerProArgTyrTrpThrAlaPheAlaGlyIleLeuLysGlnSerLeuMet 318
 QY 1072 TTCTATGGAGCCGGATACCAAGTAGAAAAAGTATTTCTCATCAATATTAGTCCAAAG 1131
 DB 319 PheTyrGlySerArgHisGlnValGluLysIleSerHisProAsnTyrAspSerLys 338
 QY 1132 ACCAAGAACAAATCAGCATTCGCTGATGAAGCTCAGAGAGCTCTGACTTTCAACAGACTA 1191
 DB 339 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnThrProLeuAlaPheAsnAspVal 358
 QY 1192 GTGAACCAAGTGTCTGTGCCCAACCCAGGCGATGCTGCAGCCAGAACAGCTCTGTCTGG 1251
 DB 359 ValLysProValCysLeuProAsnProGlyMetMetLeuAspLeuAlaGlnGluCysTrp 378
 QY 1252 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAGTGTCTCAGCGCTGCC 1311
 DB 379 IleSerGlyTrpGlyAlaThrTyrGluLysGlyLysThrSerAspValLeuAsnAlaAla 398
 QY 1312 AAGGTGCTTCTCATTGACACACAGATGCAACAGCAGAGATATCTATATGACAACTCATC 1371
 DB 399 MetValProLeuIleGluProSerLysCysAsnSerLysTyrIleTyrAsnAsnLeuLe 418
 QY 1372 ACACAGCCATGATCTGTGCGGCTTCTGCGAGGGAACCTCGATTCCTCCAGGGTGAC 1431
 DB 419 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlySerValAspSerCysGlnGlyAsp 438
 QY 1432 AGTGGAGGGCTCTGTGCTCACTTCGAAAGAACAAATATCTGGTGGTGTATAGGGGATACAAGC 1491
 DB 439 SerGlyGlyProLeuValThrLeuLysAsnGluIleTrpTrpLeuIleGlyAspThrSer 458
 QY 1492 TGGGGTCTGTGCTGCGCAAGCTTCAGACAGGAGGTACCGGAATGTGTGGTATTTC 1551
 DB 459 TrpGlySerGlyCysAlaLysAlaTyrA-gProGlyValTyrGlyAsnValThrValPhe 478

OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RF	SEQUENCE FROM N.A.
RA	Rao N.V., Rao G.N., Hoidal J.R.;
RT	"Genomic Organization of Murine Transmembrane Proteinases.";
RL	Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF479687; AAC33581.1; -
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.
DR	GO; GO:0005044; P:scavenger receptor activity; IEA.
DR	GO; GO:0004295; F:trypsin activity; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	InterPro; IPR009003; Cys_Ser_trypsin.
DR	InterPro; IPR002172; LDL_receptor_A.
DR	InterPro; IPR001254; Peptidase S1.
DR	InterPro; IPR001314; Peptidase S1A.
DR	InterPro; IPR001190; Srcr_receptor.
DR	Pfam; PF00057; ldl_recept_a, 1.
DR	Pfam; PF00089; Crypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	SMART; SM00192; LDLA_1.
DR	SMART; SM00202; SR; 1.
DR	SMART; SM00020; Tryp_SPC; 1.
DR	PROSITE; PS01209; LDLRA_1; 1.
DR	PROSITE; PS00088; LDLRA_2; 1.
DR	PROSITE; PS0287; SRCR_2; 1.
DR	PROSITE; PS0240; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
SQ	SEQUENCE 453 AA; 49505 MW; 1EE7ECD6CB3DD894 CRC64;

Alignment Scores:

Pred. No.:	8,32e-69	Length:	453
Score:	881.00	Matches:	188
Percent Similarity:	57.57%	Conservative:	63
Best Local Similarity:	43.12%	Mismatches:	159
Query Match:	27.71%	Indels:	26
DB:	11	Gaps:	12

US-09-323-597C-1 (1-1738) x Q812A6 (1-453)

Qy	316	CCCAATCCCATCGGGACAGTGTGCACCTCAAAGACTAAGAAGCACTGTGCACTC---	372
Dd	28	ProValAlaProAspGlyAspAlaValAlaAdlnIleuSerLeuLeuProLeuLys	47
Qy	373	-----ACCTTGACCCTGGGACCTTCCTCTGGAGCTGGCTGGCGCTGGCCTTA	423
Dd	48	PhePheProIleileValIleGlylleleAlaleuIleuAlaleuAlalleGlyLeu	67
Qy	424	CTCTGAAGTTTCATCGGCAGCAAGTGCTCCAACCTCTGGATAGAGTGCAGTCTCTCAGGT	483
Dd	68	GlyIleHisPhe-----AspCysSerglylys---TyrargCysHisserSerpHe	83
Qy	484	ACCTGCATCAACCCCTCTAACTGTGTGTATGGCTGTCACTGCCCGCGGGAGGAC	543
Dd	84	LysCysilleuLeuThralaargCysaspGlyValSerAspCysLysAsnalaGluasp	103
Qy	544	GAGAATCGGTGTTCGCTCTACGGACCAACTTCATCTCTCAGGTGTACTCATCTCAG	603
Dd	104	GlutYrArgCysValargvalserGlyGlnrglaalauGlnInvalPheThrila---	122
Qy	604	AGGAGTCTCTGGCACCTGTGTGTGCAAGACGACTGGAACGAGAACTACGGCGGGCGGC	663
Dd	123	---AlaAlaTrpArgThrMetCysSerAspAspTrpLysSerHisIstYrAlaLysilleAla	141
Qy	664	TGCAGGACATGGCTATAAGNAATATTTTTTACTCTTAGCCACGAATA-----	711
Dd	142	CysAlaGlnleuglyPhePro---SerItyrValserSerAspHisleuargValaspAla	160
Qy	712	GTGGATGACAGCGGATCCACCAAGCTTTATGAACACTGAACAGTGCAGGCAATGTTCAT	771

512 GlyValThrThrLysValThrGluValLeuProTfPileTyrSerLysMetGluSerGlu 531

RESULT 7

Q9BYE2 PRELIMINARY; PRT; 581 AA.
 AC Q9BYE2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Membrane-type mosaic serine protease.
 GN MSPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=21167393; PubMed=11267681;
 RA Kim D.R., Sharmin S., Inoue M., Kido H.;
 RT "Cloning and expression of novel mosaic serine proteases with and
 RT without a transmembrane domain from human lung";
 RL Biochim. Biophys. Acta 1518:204-209 (2001).
 DR EMBL; AB048796; BAB39741.1; -;
 DR HSSP; P00763; LDPO.
 DR MEROPS; S01.087; -;
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR002172; LDL receptor A.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase S1a.
 DR InterPro; IPR001190; Srtr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLa; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0287; SRCK_2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA44 CRC64;

Alignment Scores:

Pred. No.: 581 Length: 581
 Score: 736.50 Matches: 177
 Percent Similarity: 46.67% Conservative: 82
 Best Local Similarity: 31.89% Mismatches: 197
 Query Match: 23.17% Indels: 99
 DB: 4 Gaps: 17

US-09-323-597C-1 (1-1738) x Q9BYE2 (1-581)

QY	124	TCAGGTGTCACCACTGCTTGGACCTTACTATGAAACCATGGATACCAACCGGAAC	183
DB	27	AlaGlyThrProGlyArgala-----SerProAlaGln	38
QY	184	CCCTATCCCGACAGCCCACTGGGCCCACT---GHCTACGAGGTGATCCGGCTCAG	240
DB	39	AlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgAlaSerProAlaGln	58
QY	241	TACTACCGTCCCGGTGCC-----CAGTACGCCCGGAGGTG	279
DB	59	AlaSerProAlaGlyThrProGlyArgAlaSerProGlyArgAlaSerProAlaGln	78
QY	280	CTGACGACGGCT---TCCAAACCGGTGCTCCACGACGCCAAATCCCATCCGGG---	333
DB	79	AlaSerProAlaGlnAlaSerProAlaGlnAlaSerLeuSerArgSerSerGlyArg	98

QY	334	-----ACAGTGTGCACTCAAGAGCTAAAGAAAGCACTGTGCATC	372
DB	99	SerSerSerAlaArgSerAlaSerValThrThrSerProThrArgValTyrLeuValArg	118
QY	373	ACCTTGACCCCTGGGACCTTC-----	393
DB	119	AlaThrProValGlyAlaValProIleArgSerSerProAlaArgSerAlaProAlaThr	138
QY	393	-----	393
DB	139	ArgAlaThrArgGluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGln	158
QY	394	-----CTCGTGGAGCT-----GCCTGGCGCTGGCTACTC	426
DB	159	LysGlnLeuProLeuIleGlyCysValLeuLeuLeuLeuAlaLeuValValSerLeuIle	178
QY	427	-----TGAAGTTCATGGCAGCAAGTCTCAACTCTGGGATA--GAG	468
DB	179	IleLeuPheGlnPheTrpGln-----GlyHisThrGlyIleArgHis	192
QY	469	TGCAGCTCTCAGGTACTGTCATCAACCCCTTAACGTGTGTGATGCGTGTCACTGC	528
DB	193	LysGlnArgGluSerCysProLysHisAlaValArgCysAspGlyValValAspCys	212
QY	529	CCCGCGGGGAGGAGCAATCGGTGTGCTCTACGACCAAACTTCATCTTCAG	588
DB	213	LysLeuLysSerAspGluLeuGlyCysValArgPheAspTrpAspLysSerLeuLeuLys	232
QY	589	GTGTACTCATCTCAGAGAAAGTCTCTGTCGCAACCTGTGTCACCAAGCACTGGAAGCAAC	648
DB	233	IleTyrSerGlySerSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSer	252
QY	649	TACGGCGGGCGGCTCAGGACATCGGTATAGAATAATTTTACTTAGCCAA---	705
DB	253	TyrSerGluLysThrCysArgGlnLeuGlyPheGluSerAlaHisArgThrGluVal	272
QY	706	GGAATAGTGTAGCAGCGGATCCACAGCTTTATGAACCTTGAACCAAGTCCGCGCAAT	765
DB	273	AlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThr-----	289
QY	766	GTCGATATCTATAAAACTGTACACAGTATCCCTGCTTCTTCTTCAAAGCACTGTTCT	825
DB	290	-----IleGlnGluSerLeuHisArgSerHisCysProSerGlnArgTyrIleSer	306
QY	826	TTACGCTGTATAGCTCGGGGTCACTTGAACCTCAAGCGCCAGACGAGTATGTGGC	885
DB	307	LeuGlnCysSerHisCysGlyLeuArg-----AlaMetThrGlyArgIleValGly	323
QY	886	GGCGAGAGCGGCTCCCGGGGCGCTGCCCTGCGAGTCACTGCGTCCAGTCCAGAACGTC	945
DB	324	GlyAlaLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThrThr	343
QY	946	CACGTGTGGGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCGCCCACTGGCTG	1005
DB	344	HisIleCysGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhe	363
QY	1006	-----GAAAAACCTCTTAACTCCATTCGCACTGGCACTTGGCGGCACTTGGCGGAT	1053
DB	364	PheValThrArgGluLysValLeuGluGly-----TrpLysValTyrAlaGlyThr	380
QY	1054	TTGAGACATCTTTCATGTTCTATGAGCGGATACCAAGTAGAGAAAAGTATTTCTCAT	1113
DB	381	SerAsnLeuHisGlnLeuProGluAlaAla-----SerIleAlaGluIleIleAsn	398
QY	1114	CCAATATGACTCCAAGCAACCAATGACATTCGCTGATGAAGCTCAGAGCTCAGAAGCT	1173
DB	399	SerAsnTyrThrAspGluGluAspAspTyrAspIleAlaLeuMetArgLeuSerLysPro	418
QY	1174	CTGACTTTCAACGACCTAGTGAACACGATGTGTCTGCCCAACCCAGGATGATGCTGAC	1233
DB	419	LeuThrLeuSerAlaHisIleHisProAlaCysLeuProMetHisGlyGlnThrPheSer	438
QY	1234	CCAGAACAGCTCTCTCGATTTCGCGGTGGGGGGGCCCGAGGAG---AAAGGAAGACC	1290

```

Db 439 LeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluThrAspLysThr 458
QY 1291 TCAGAGAGTGTGAAGCTGCGCAGAGTCTTCTCATTGACAGACAGAGATGCAACACAGA 1350
Db 459 SerProPheLeuArgGluValGlnValAsnLeuLeuLeuLeuLeuLeuLeuLeuLeu 478
QY 1351 TATGTCTATGACAACTGATACACAGCAGGCGATGCTGTGCGGCTTCTGCGAGGGGAAAC 1410
Db 479 LeuValTyAspSerTyLeuThrProArgMetMetCysAlaGlyAspLeuHisGlyGly 498
QY 1411 GTGATCTTCCAGAGGTGACATGAGGCGCTCTGCTCATTCTCGAAGACAATATCTCG 1470
Db 499 ArgAspSerCysGlnGlyAspSerGlyGlyProLeuValCysGluGlnAsnAsnArgTrp 518
QY 1471 TGGCTATGAGGATACAACTGCGGTCTGCTGCTGCGCAGGCTTACAGACAGAGAGTG 1530
Db 519 TyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyVal 538
QY 1531 TAGCGGAATGTGATGATTACAGGACTGCGATTTATCGACAATG 1575
Db 539 TyrThrLysValThrGluValLeuProTrpIleTySerLysMet 553

RESULT 8
QSCFE0 PRELIMINARY; PRT; 471 AA.
AC QSCFE0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to mosaic serine protease (Fragment).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAUN=C57BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042878; AAH42878.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Srcr receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00330; SRCR; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tlyp_SPC; 1.
DR PROSITE; PS00267; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
FT NON TER.
SQ SEQUENCE 471 AA; 52535 MW; ED58CFB67C3BCC4 CRC64;

```

Alignment Scores:

```

Pred. No.: 1,45e-55 Length: 471
Score: 731.00 Matches: 166
Percent Similarity: 51.50% Conservative: 75
Best Local Similarity: 35.47% Mismatches: 183
Query Match: 22.99% Indels: 44
DB: 11 Gaps: 13

```

```

US-09-323-597C-1 (1-1738) x Q8CFE0 (1-471)
QY 250 TCCCCGGTGCCCCAGATAGCCCGCGAGGGTCTCGACGAGGCTTCCACCCCGTCTGCTGC 309
Db 19 SerProThrArgValTyLeuValArgAlaThrProValGlyAlaValProIleArgAla 38
QY 310 AGCAGCCCAATATCCCATCCCGGACAGTGTGCACC----- 345
Db 39 SerProAlaArgSerAlaProAlaThrArgAlaThrArgGluSerProGlyLeuSerPhe 58
QY 346 -----TCAAGACTAAGAAAGCACTG-----TGCATCACCTTG 378
Db 59 ProLysPheSerTrpGlnGluThrGlnArgGlnLeuProLeuLeuLeuLeuLeuLeu 78
QY 379 ACCCTGGGACCTTCTCGTGGGAGCTGCGCTGGCGCTGCTACTCTGGAAGTTCATG 438
Db 79 -----LeuLeuSerLeuValLeuLeuLeuLeuLeuLeuPheTyTrp 93
QY 439 GGCAGCAAGTGTCCAACTCTGGGATAGAGTGC---GACTCTCTAGGTACCTGCATCAAC 495
Db 94 Arg-----GlyHisThrGlyIleLysTyLysGluProLeuGluSerCysProIle 110
QY 496 CCTCTAACTGGTGTGGGTGTACACATGCCCCGCGGAGGAGACAGAAATCGGTGT 555
Db 111 HisAlaValArgCysAspGlyValValAspCysLysMetLysSerAspGluLeuGlyCys 130
QY 556 GTTCGCTCTACGGACCAAACTTCTCTCAGGTGTACTCATCTCAGAGGAAGTCTGTGG 615
Db 131 ValArgPheAspTrpAspLysSerLeuLeuLysValTySerGlySerGlyGluTrp 150
QY 616 CACCCTGTGTGCGCAAGACGACTGGACAGAGACTACGGCGGGCGGCTGCGAGGACATG 675
Db 151 LeuProValCysSerSerSerTrpAsnAspThrAspSerLysArgThrCysGlnGluLeu 170
QY 676 GGTATAAGAAATAATTTTACTCTAGCCAAGAAATAGTGGATGACAGCGGATCCACGAC 735
Db 171 GlyPhe---AspSerAlaTyArgThrThrGluValAlaHisArgAspIleThrSerSer 189
QY 736 TTTATGAACCTGACACAGAGTCCCGGCAATGTGATATCTATAAAACTGTACACAGT 795
Db 190 PheLeu-----LeuSerGluTyArgThrThrIleGlnGluSerLeuTyArgSer 206
QY 796 GATGCTCTCTTCAAAAGCAGTGTGTTTCTTACCTGTATAGCTCGGGGTCAACTTG 855
Db 207 Gln---CysProSerArgArgTyValSerLeuGlnCysSerHisCysGlyLeuArg--- 224
QY 856 AACTCAAGCCGCCAGACAGCAGATGTGGCGGCGAGAGCGGCTCCCGGGGCGCTGGCCC 915
Db 225 -----AlaMetThrGlyArgIleValGlyGlyAlaLeuThrSerGluSerLysTrpPro 242
QY 916 TGGCAGGTGACGCTGACGTCCAGAACGTCACAGTGTGCGGAGGCTCCATCATCACCCC 975
Db 243 TrpGlnValSerLeuHisPheGlyThrThrHisIleCysGlyGlyThrIleLeuLeuAsp 262
QY 976 GAGTGGATCGTGACAGCGCCCACTGCGTG-----GAAAAACCTCTTAAACAAT 1023
Db 263 GlnTrpValLeuThrAlaAlaHisCysPheValThrArgGluLysLeuGluGly 282
QY 1024 CCATGCGCATTCGACGGCATTTCCGGGATTTTGAGACAACTCTTTCATGTTCTATGAGCC 1083
Db 283 -----TrpLysValTyAlaGlyThrSerAsnLeuHisGlnLeuProGluAla 299
QY 1084 GGATACCAAGTAGAAAAAGTATTCTCATCAAAATATGACTCCCAAGACCAAGAACAAAT 1143
Db 300 -----SerIleSerGlnIleIleAsnGlyAsnTyThrAspGluLeuAspAspTy 317
QY 1144 GACATTGGCTGATGACAGCTCGACAGCTCTGACTTTCAACGACCTAGTGAACCACTG 1203
Db 318 AspIleAlaLeuIleArgLeuSerLysProLeuThrLeuSerAlaHisIleProAla 337
QY 1204 TGCTGCGCCCAACCCAGGCAATGATGTCGCGAGCCAGAACAGCTCTGCTGATTTCCGGGTG 1263

```

```

Db      338 CysLeuProMetHisGlyGlnThrPheGlyLeuAsnGluThrCysTrpIleThrGlyPhe 357
QY      1264 GGGGCCACCGAGAGAGAGG---AAGACTCAGAAAGTCTGACGCTCCCAAGTGCTT 1320
Db      358 GlyLeuThrLysGlnThrAspGluLysThrSerProPheLeuArgGluValGlnValAsn 377
QY      1321 CTCATTGACACACAGAGATGCAACACAGCATATGCTATGACAACTGATCACACAGCC 1380
Db      378 LeuIleAspPheLysCysAsnAspTyrLeuValTyrAspSerTyrLeuThrProArg 397
QY      1381 ATGATCTGTGCGGCTTCTCGAGGGAGAACGTCGATCTTTCAGGGTGACAGTGAGGG 1440
Db      398 MetMetCysAlaGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGly 417
QY      1441 CTTCTGCTCCTTCTGGAAGAACATATCTGCTGCTGATAGGAGGATACAGCTGGGTTCT 1500
Db      418 ProLeuValCysGluGlnAsnAsnArgTyrPheLeuAlaGlyValThrSerTrpGlyThr 437
QY      1501 GGCTGTGCGCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTCACGAGCTGG 1560
Db      438 GlyCysGlyGlnLysAsnLysProGlyValTyrThrLysValThrGluValLeuProTrp 457
QY      1561 ATTATTCGACAATGAGGCGACAC 1584
Db      458 IleTyrArgLysMetGluSerGlu 465

RESULT 9
Q86YM4
ID      Q86YM4; PRELIMINARY; PRT; 558 AA.
AC      Q86YM4;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Transmembrane protease serine 6.
GN      TMPSR6.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Park T.J., Park W.J.;
RT      "Homo sapiens transmembrane protease, serine 6 (TMPSR6) mRNA.";
RL      Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY190317; AAO38062.1; -;
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0004263; F:chymotrypsin activity; IEA.
DR      GO; GO:0008233; F:peptidase activity; IEA.
DR      GO; GO:0005044; F:scavenger receptor activity; IEA.
DR      GO; GO:0004295; F:trypsin activity; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      InterPro; IPR009003; Cys_Ser_trypsin.
DR      InterPro; IPR001254; Peptidase S1.
DR      InterPro; IPR001314; Peptidase S1A.
DR      Pfam; PF00089; Srcr_receptor.
DR      InterPro; IPR001190; Srcr_receptor.
DR      PRINTS; PR00722; trypsin; 1.
DR      SMART; SM00202; SR; 1.
DR      SMART; SM00020; Tryp_SPC; 1.
DR      PROSITE; PS0287; SRCR_2; 1.
DR      PROSITE; PS0240; TRYPSIN_DOM; 1.
DR      PROSITE; PS00134; TRYPSIN_HIS; 1.
DR      PROSITE; PS00135; TRYPSIN_SER; 1.
KW      Protease.
SQ      SEQUENCE 558 AA; 60432 MW; 4AC817FCD70D7017 CRC64;

Alignment Scores:
Pred. No.: 4,65e-55 Length: 558
Score: 725.50 Matches: 174
Percent Similarity: 45.39% Conservative: 82
Best Local Similarity: 30.85% Mismatches: 195
Query Match: 22.82% Indels: 113
DB: 4 Gaps: 16

```

```

US-09-323-597C-1 (1-1738) x Q86YM4 (1-558)
QY      124 TCAGGTCACACACAGCTATTGGACCTTACTATGAAACCATGATACCAACCGAAGAAC 183
Db      27 AlaGlyThrProGlyArgAla-----SerProAlaGln 38
QY      184 CCCTATCCCGACACGCCCACTGTGTGCCCACT---GTCTACGAGGTGCATCCGCTCAG 240
Db      39 AlaserProAlaGlnAlaserProAlaGlyThrProGlyArgAlaserProAlaGln 58
QY      241 TACTACCGCTCCCGCTGCC---CAGTACGCCCGGAGGCTC 279
Db      59 AlaserProAlaGlyThrProGlyArgAlaserProGlyArgAlaserProAlaGln 78
QY      280 CTGAGCGAGGT---TCAACCCGCTGCTGCGACGCCCAAAATCCCACTCCGG--- 333
Db      79 AlaserProAlaArgAlaserProAlaLeuAlaserLeuSerArgSerSerGlyArg 98
QY      334 -----ACAGTGTGCACTCAAAGACTAAAGACACTGTGCATC 372
Db      99 SerSerAlaArgSerAlaserValThrThrSerProThrArgValTyrLeuValArg 118
QY      373 ACCTTGACCTGGGACCTTC----- 393
Db      119 AlaThrProValGlyAlaValProIleArgSerProAlaArgSerAlaProAlaThr 138
QY      393 ----- 393
Db      139 ArgAlaThrArgGluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGln 158
QY      394 -----CTGCTGGAGCTGCCCTGCC----- 414
Db      159 LysGlnLeuProLeuIleGlyCysValLeuLeuLeuAlaLeuValValSerLeuIle 178
QY      415 -----GCTGGCTACTCTCGAAGTTCATCGGACGACAG 447
Db      179 IleLeuPheGlnPheTrpGlnGlyTyrThrGlyIleArgTyrLysGluGlnArgGluSer 198
QY      448 TGCTCCCAACTCTGGGATAGAGTGGGACTCCTCAGGTACTGTCATCAACCCCTCTAACTGG 507
Db      199 CysProGluHisAlaValArg----- 205
QY      508 TGTGATGGCGTGTCACTCCCGCGGGGAGGAGACAGAAATCGGTGTGTGGCTCTAC 567
Db      206 ArgAspGlyValValAspCysLysLeuLysSerAspGluLeuGlyCysValArgPheAsp 225
QY      568 GGACCAAACTTCTCTCAGGTGTCTCATCTCAGAGGAAGTCTCTGACCCCTGTGTC 627
Db      226 TrpAspLysSerLeuLeuLysIleTyrSerGlySerSerHisGlnTrpLeuProIleCys 245
QY      628 CAAGACGACTGGAACGAGAACTACGGCGCGCGCTGCGAGGACATGGGCTATAAGAAT 687
Db      246 SerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSer 265
QY      688 AATTTTACTCTAGCCAA---GGAATAGTGTGATGACAGCGGATCCACAGCTTTATGAAA 744
Db      266 AlaHisArgThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArg 285
QY      745 CTGAACACAAAGTGCCTGCAATGTCTATCTATAAAACTGTACACAGATGATGCTGT 804
Db      286 TyrAsnSerThr-----IleGlnGluSerLeuHisArgSerGlu---Cys 299
QY      805 TCTTCAAAAGCAGTGGTTTCTTTACGCTGTATAGCTCGCGGTCACTTGAACCTCAAGC 864
Db      300 ProSerGlnArgTyrIleSerLeuGlnCysSerHisCysGlyLeuArg-----Ala 316
QY      865 CGCCAGACGAGGATTTGGCGCGGAGAGCGCGCTCCCGGGGGCTCGCCCTGCGAGCTC 924
Db      317 MetThrGlyArgIleValGlyAlaLeuAlaserAspSerLysTyrProTrpGlnVal 336
QY      925 AGCTGCACTCCAGAACGTCCAGGTGTGCGAGGCTCCATCATCCCCCGAGTGGATC 984

```



```
Db 337 SerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpVal 356
Qy 985 GTGACGCGCCCTCGTGC-----GAAAGAACTCTTAACAATCCATGCGCAT 1032
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 357 LeuThrAlaAlaHisCysPhePheValThrArgGluLysValLeuGluGly----- 373
Qy 1033 TGGACGCGCATTTGGGGGATTTGAGACAATCTTTCATGTTCTTCTATGAGCGCGANACAA 1092
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 374 TrpLysValThrAlaGlyThrSerAsnLeuHisGlnLeuProGlnAlaAla-----Ser 391
Qy 1093 GTAGAAAAGTATTTCTCATCAAAATATGACTCCAGACCAAGACCAAGAAATGACATTCGCG 1152
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 392 IleAlaGluIleIleAsnSerAsnThrThrAspGluGluAspAspTyrAspIleAla 411
Qy 1153 CTGATGAAGCTGACAGACCTCTGACTTTCACAGCCTAGTGAACCAAGTGTGTCTGCC 1212
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 412 LeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleHisProAlaCysLeuPro 431
Qy 1213 AACCCAGGCGATGCTGCAGCCAGACAGCTCTGCTGGATTTCCGGGTGGGGCCACC 1272
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 432 MethisGlyGlnThrPheSerLeuAsnGluThrCysTrpIleHmrGlyPheGlyLysThr 451
Qy 1273 GAGGAG---AAGGGAAGACCTCAGAAAGTGTGTAACGCTGCCAAGTGTCTTCTCATTCGAG 1329
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 452 ArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuIleAsp 471
Qy 1330 ACACAGAGATGCACACAGATATGCTATGACACCACTGATCATCACACAGCCATGATCTGT 1389
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 472 PheLysCysAsnAspTyrLeuValThrAspSerTyrLeuThrProArgMetMetCys 491
Qy 1390 GCCGGCTTCCTGCAGGGGAAAGCTCGATTCTTCCAGGGTGCACAGTGGAGGCTCTGCTGC 1449
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 492 AlaGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyProLeuVal 511
Qy 1450 ACTTCGAAGAACATATCTGTGCTGATAGGGAATACAGCTGGGCTTCTGCTGTGCC 1509
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 512 CysGluGlnAsnAsnArgTrpTrpLeuAlaGlyValThrSerTrpGlyThrGlyCysGly 531
Qy 1510 AAGCTTACAGACACAGAGTGTGCGGAATGCTGATGCTATTCAGGACTGGATTTATCGA 1569
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 532 GlnArgAsnLysProGlyValThrLysValThrLysValLeuValLeuProTrpIleTyrSer 551
Qy 1570 CAANTGAGGCA 1581
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 552 LysMetGluSer 555

RESULT 10
Q96B86 PRELIMINARY; PRT; 405 AA.
AC Q96B86;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC 1-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -.
DR HSSP; P00761; 1AN1.
DR MEROPS; S01.034; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004283; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
DR InterPro; IPR009003; Cys_Ser trypsin.
DR InterPro; IPR002172; LDL_receptor A.
DR InterPro; IPR001254; Peptidase_S1_
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM0020; Tryp_SPC; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
FT NON TER 1 1
SQ SEQUENCE 405 AA; 44474 MW; 951ACD52D9D48E04 CRC64;

Alignment Scores: 8.97e-51 Length: 405
Pred. No.: 676.50 Matches: 150
Score: 676.50
Percent Similarity: 53.91% Conservative: 57
Best Local Similarity: 39.06% Mismatches: 128
Query Match: 21.28% Indels: 49
DB: 4 Gaps: 13

US-09-323-597C-1 (1-1738) x Q96B86 (1-405)
Qy 508 TGTGATGGGTGTCACACCTGCCCGGGGAGGAGAGAAATCGTGTTCGCTCTAC 567
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 45 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 64
Qy 568 -----GGACCA-----AACTTCATCTTCAGGTGTAC 594
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 84
Qy 595 TCATCTCAGAGGAAGCTCGGCACCTGTGTGTCACAGACGACATGGAACGAGAACTACGGG 654
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 104
Qy 655 CGGCGCGCTCGAGGACATGGGCTAT-----AAGATAATATTTTACTTACGCCAA--- 705
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle 124
Qy 706 -----GGAATAGTGGATGACAGCGGATCCACCGACTTATGAACCTG 747
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 144
Qy 748 AACACAAGTCCCGGCAATGTCGATATCTATAAAAAAATGTACACAGTGTCTTCT 807
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 ArgAsnSerSerGly-----ProCysLeu 152
Qy 808 TCNAAAGCAGTGGTTTCTTTTACCTGTATAGCTGGGGGTCAACTTGAACCTCAAGCCGC 867
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 153 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 169
Qy 868 CAGAGCAGCATTTGTGGCGCGCAGAGCGGCTCCCGGGGGCTGGCCCTGGCAGGTGAGC 927
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 170 ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer 189
Qy 928 CTGCAGCTCCAGAACTCCAGTGGCGAGGCTCCATCATCATCCCCCGAGTGGATCGTG 987
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 209
Qy 988 ACAGCGCGCCACTGGCTGGGAAAAAOCCTCTTAACAATCCATGGCATGGACGGCATTTGCG 1047
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 210 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 228
Qy 1048 GGGATTTTGAGA---CAATCTTTTCTATGTTCTATGGAGCGGATACCAAGTAGAAAAAGTG 1104
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 229 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 244
```



```
QY 1122 TGACTCCAGACCAAGACATGATTCGCTGATGAGCTGAGAGCTCTGACTTT 1181
Db 290 rSerAlaGlnAenHisAspTyrAspValAlaLeuGlnLeuArgThrProIleAenPh 310
QY 1182 CAACGACCTAGTGAACACAGTGTCTGCCCCAACCCAGGATGATGCTGACGCCAGAAC 1241
Db 310 eSerAspThrValSerAlaValCysLeuProAlaLysGlnGlnHisPheProGlnGlySe 330
QY 1242 GCTCTGCTGATTTCCGGTCCGGGCGCCACCGAG---GAGAAAGGAGACCTCAGAGT 1298
Db 330 rGlnCysrPValSerGlyrPGLyHisThrAspProSerHisThrHisSerAspTh 350
QY 1299 GCTGAACGCTCCAGGTGCTTCATTTGACAGACAGAGATGACACAGCAGATATGCTA 1358
Db 350 rLeuGlnAspThrMetValProLeuLeuSerThrAspLeuCysAsnSerSerCysMetTy 370
QY 1359 TGACACCTGATACACAGCCATGATCTGCGCGGCTTCCTGAGGGGAACTGCGATTC 1418
Db 370 rSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyrLeuAspGlyArgAlaAspAl 390
QY 1419 TTGCGAGGCTCACAGTGGAGGCTCTGCTCACCTCGAGAACATATCTGGTGGCTGAT 1478
Db 390 aCysGlnGlyAspSerGlyGlyProLeuValCysProSerGlyAspThrTrpHisLeuVa 410
QY 1479 AGGGATACAGCTGGGGTCTGCTGTGCGAAAGCTTACAGACGAGAGTGTACGGGAA 1538
Db 410 lGlyValValSerTrpGlyArgGlyCysAlaGluProAsnArgProGlyValTyAlaLy 430
QY 1539 TGTGATGTTATCCAGCACTGGATTAT 1566
Db 430 sValAlaGluPheLeuAspTrpIleHis 439

RESULT 12
Q8CDRO PRELIMINARY; PRT; 455 AA.
AC Q8CDRO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transmembrane protease.
GN TPFRS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).";
DR EMBL; AK029714; BAC26577.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00089; trypsin; I.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 455 AA; 49669 MW; BE22EB2E7503C74B CRC64;
```

```
Alignment Scores:
Pred. No.: 1,06e-49 Length: 455
Score: 664.50 Matches: 145
Percent Similarity: 50.97% Conservative: 66
Best Local Similarity: 35.02% Mismatches: 176
Query Match: 20.90% Indels: 27
DB: 11 Gaps: 9

US-09-323-597C-1 (1-1738) x Q8CDRO (1-455)
QY 367 TGCATCACCTTGACCTCGGGGACCTTCTCTGCTGGAGCTCGCTGGCGCTGGCTACTC 426
Db 49 CysValIleLeuGlyValLeuGlyLeuLeuAlaGlyAlaGlyIleAlaSerTrpLeuLeu 68
QY 427 TGAAGTTCTAAGGCGAGCAAGTCTCCAATCTTGGCATAGAGTTCGAGTCTCTCAGTACC 486
Db 69 ValLeuTyLeuTrpProAlaSerProSer-----IleSerGlyThr 83
QY 487 TGCATCACCCCTTAACCTGGTGTGATGGCGTCTCACACTGCGCGCGGGAGGAGGAG 546
Db 84 LeuGlnGluGlu-----GluMetThrLeuAsnCysProGlyValSerArgGlu 99
QY 547 AATCGGTGTGTT-----CGCCTCTACGACCAAACTTC 579
Db 100 GluGluLeuLeuProSerLeuProLysThrValSerPheArgIleAsnGlyGluAspLeu 119
QY 580 ATCCTTCAGTGTACTCTCTCAGAGAAAGTCTCGGACCCCTGTGTGCCAAGAGACTGG 639
Db 120 LeuLeuGlnValGlnValAlaArgProAspTrpLeuLeuValCysHisGluGlyTrp 139
QY 640 AACGAGAACTACGGCGGGCGGCTCAGGAGATGGGCTAT---AAGATAATTTTTC 696
Db 140 SerProAlaLeuGlyMetHisIleCysLysSerLeuGlyHisIleArgLeuThrGlnHis 159
QY 697 TCTAGCCAAAGATAGTGGATGAC-----AGCGATCCACAGCTTTATGAACGAAAC 750
Db 160 LysAlaValAsnLeuSerAspIleLysLeuAsnArgSerGlnGlnPheAlaGlnLeuSer 179
QY 751 ACAAGTGGCGCAATCTCGATATCTATAAAACTTACCACAGTGCATCGCTTCTTCA 810
Db 180 AlaArgProGlyGlyLeu---ValGluGluSerTrpLysProSerSerAlaSerCysProSer 198
QY 811 AAGCAGTGGTTTCTTTACGTGTATAGCTCGGGGTCAACTTGAACCTCAAGCCGCCAG 870
Db 199 GlyArgIleValSerLeuLysCysSerGluCysGlyAlaArg-----ProLeuAla 215
QY 871 AGCAGGATTCGGCGGGCGAGAGCGGCTCCCGGGGCGCTCGGCTGGCAGGTCAGCCTG 930
Db 216 SerArgIleValGlyGlnAlaValAlaSerGlyArgTrpProTrpGlnAlaSerVal 235
QY 931 CACGTCCAAAGTCCACGTGTCGAGGCTCCATCATCACCCCGAGTGGATCGTGACA 990
Db 236 MetLeuGlySerArgHisThrCysGlyAlaSerValLeuAlaProHisTrpValValThr 255
QY 991 GCGGCCACTGCGTGGAATA---CCTTTAACAATCCATGCGATTGGACGGCATTTGGG 1047
Db 256 AlaAlaHisCysMetTySerPheArgLeuSerArgLeuSerSerTrpArgValHisAla 275
QY 1048 GGGATTTTCAGACAATCTTTTCATGTTCTTATGCGCGGATCCACAGTACAAAAGTATT 1107
Db 276 GlyLeuValSerHisGlyAlaValArgGlnHisGlnGlyThrMetValGluLysIleIle 295
QY 1108 TCTCATCCAAATATGATCTCCAAAGACCAAGAACATGACATTCGCGTGTAGAGTGCAG 1167
Db 296 ProHisProLeuTySerAlaGlnAsnHisAspTyArgValAlaLeuLeuGlnLeuArg 315
QY 1168 AAGCCTCTGACTTTCAACGACCTAGTGAACACGAGTGTCTGCCCAACCCAGGCATGATG 1227
Db 316 ThrProIleAsnPheSerAspThrValGlyAlaValCysLeuProAlaLysGluGlnHis 335
QY 1228 CTCGAGCCCAACACAGCTCTCTGTGGATTTCGGGTGGGGGGCCACCGAG---GAGAAAGG 1284
```

```
Db 336 PheProTyrGlySerClnCysTrpValSerGlyTyrGlyHisThrAspProSerHisThr 355
QY 1285 ANACCTCAGAGTCTGCAAGCTGCCAAGGTGCTTCTCATTCAGACACAGAGATGCAAC 1344
Db 356 HisSerSerAspThrLeuGlnAspThrMetValProLeuLeuSerThrTyrLeuCysAsn 375
QY 1345 AGCAGATATGCTATGACAACTGATCACCAGCAGCATGATGTGCGGCTTCCTCGAG 1404
Db 376 SerSerCysMetTyrSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyTyrLeuAsp 395
QY 1405 GGAACCTGATCTTCCAGAGTGCAGTGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1464
Db 396 GlyArgAlaAspAlaCysGlnGlyAspSerGlyGlyProLeuValCysProSerGlyAsp 415
QY 1465 ATCTGCTGCTGATAGGAGTACCAAGCTGGGCTTCTGGCTGTGCCAAAGCTTACAGACCA 1524
Db 416 ThrTrpHisLeuValGlyValValSerTrpGlyArgGlyCysAlaGlnProAsnArgPro 435
QY 1525 GGAGTGCAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1566
Db 436 GlyValTyrAlaLysValAlaGluPheLeuAspTrpIleHis 449

RESULT 13
Q8CJ16 PRELIMINARY; PRT; 371 AA.
AC Q8CJ16;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-NOV-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adrenal mitochondrial protease short variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEOH;
RA Omer S.; Bicknell A.B.; Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF537099; AAN06758.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srrc_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYP_SIN_DOM; 1.
DR PROSITE; PS00134; TRYP_SIN_HIS; 1.
DR PROSITE; PS00135; TRYP_SIN_SER; 1.
KW Protease.
SQ SEQUENCE 371 AA; 40694 MW; 89A6408D19A1FE26 CRC64;
```

```
Alignment Scores:
Pred. No.: 1,51e-49 Length: 371
Score: 662.50 Matches: 133
Percent Similarity: 53.82% Conservative: 57
Best Local Similarity: 37.68% Mismatches: 148
Query Match: 20.84% Indels: 15
DB: 11 Gaps: 6
```

US-09-323-597C-1 (1-1738) x Q8CJ16 (1-371)

QY 532 GCGGGGAGGACGAGATCGGTGGTTCCTCTACGGACCAAACTTCATCCTTCAGGTG 591

```
Db 20 GlyAspGlnGlnGlnProIleSerPheArgIleAsnGlyGluAspLeuLeuGluVal 39
QY 592 TACTCATCTCAGAGAAAGTCTCGCACCTGTGTGCCAAGACGACGCTGAAACGAGAACTAC 651
Db 40 GlnValArgAlaArgProAspTrpLeuLeuValCysHisGlnGlyTyrAsnProAlaLeu 59
QY 652 GGGCGGGCGGCTCAGGAGCATGGCTATAAGATAATTTTACTCTAGCCAA----- 705
Db 60 GlyMetHisIleCysGlnSerLeuGlyTyr-----PheArgLeuThrGlnHisLys 76
QY 706 -----GGAATAGTGATGACAGCGGATCCACCACTTTATGAACCTGAACACACA 753
Db 77 AlaValAsnLeuSerAspIleLysLeuAsnArgSerGlnGluPheAlaGlnLeuSerAla 96
QY 754 AGTCCCGCAATGTCGATATCTATAAAACTGTACACAGTGTGCTGCTCTCTCAAAA 813
Db 97 ArgProGlySerLeu---ValGluGluAlaTrpGlnProSerThrAsnCysProSerGly 115
QY 814 GCAGTGGTCTTTTACCTGTATAGCTCGGGGTCACTTGAACCTCAAGCCGCGCAGAGC 873
Db 116 ArgIleValSerLeuLysCysSerGluCysGlyAlaArg-----ProLeuAlaSer 132
QY 874 AGATTGTGGCGCGCAGAGCGGCTCCCGGGGCTGGCCCTGGCAGGTGAGCTGCGAC 933
Db 133 ArgIleValGlyGlnAlaValAlaSerGlyArgTrpProTrpGlnAlaSerValMet 152
QY 934 GTCCAGAACGTCCAGGTGCGGAGGCTCATCATCACCCGCGAGTGTGATGTCAGACGC 993
Db 153 LeuGlySerArgHisThrCysGlyGlySerValLeuAlaProTyrTrpValValThrAla 172
QY 994 GCCCACTGCTGGGAAAAA---CCTTTAACAACTCATGGCATTCGACGGCATTTGCGGG 1050
Db 173 AlaHisCysMetTyrSerPheArgLeuSerArgLeuSerSerTrpArgValHisAlaGly 192
QY 1051 ATTTTGAGACAACTCTTTCATGTTCTATGGAGCGGATACCAAGTAGAAAAAGTATTCT 1110
Db 193 LeuValSerHisSerAlaValArgGlnHisGlnGlyThrMetValGluLysIleIlePro 212
QY 1111 CATCAAAATTATGATCCAGACCAAGAACAAATGACATTGCGTGTGATGAGTGCAGAG 1170
Db 213 HisProLeuTyrSerAlaGlnAsnHisAspTyrAspValAlaLeuLeuGlnLeuArgThr 232
QY 1171 CCTCTGACTTTCAACGACCTAGTGAACACCTGTGTCGCCCAACCCAGGCATCATGCTG 1230
Db 233 ProIleAsnPheSerAspThrValSerAlaValCysLeuProAlaLysGluGlnHisPhe 252
QY 1231 CAGCCAGAACAGCTCTGCTGGATTTCGGGTGGGGGGGCCACCCGAG---GAGAAAGGGAAG 1287
Db 253 ProGlnGlySerGlnCysTrpValSerGlyTyrGlyHisThrAspProSerHisThrHis 272
QY 1288 ACCTCAGAGTGTGTAACCGTGCACAGGTGCTTCTCATTTGAGACACAGATGCAACAGC 1347
Db 273 SerSerAspThrLeuGlnAspThrMetValProLeuLeuSerThrAspLeuCysAsnSer 292
QY 1348 AGATATGCTCTATGACAACTGATCACCAGCCATGATCTGTCCGGCTTCTCTGCGAGGG 1407
Db 293 SerCysMetTyrSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyTyrLeuAspGly 312
QY 1408 AACGTGATTTGCGCAGGGTGCAGTGGAGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1467
Db 313 ArgAlaAspAlaCysGlnGlyAspSerGlyGlyProLeuValCysProSerGlyAspThr 332
QY 1468 TGGTGGCTGATAGGGGATACCAAGCTGGGTTCCTGGCTGCTGCCAAAGCTTACAGACGAG 1527
Db 333 TrpHisLeuValGlyValValSerTrpGlyArgGlyCysAlaGluProAsnArgProGly 352
QY 1528 GTCTACGGGAATGTGATGATGATTCACGGAGCTGATTTAT 1566
Db 353 ValTyrAlaLysValAlaGluPheLeuAspTrpIleHis 365

RESULT 14
Q7Z280
```

```

ID Q7Z280 PRELIMINARY; PRT; 326 AA.
AC Q7Z280;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI-dZ69G10.3 (Novel protein similar to human transmembrane protease,
DE serine 3 (TPRSS3)) (Fragment).
DE SI-dZ69G10.3.
GN Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL672083; CAD61105.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0004263; F: chymotrypsin activity; IEA.
DR GO; GO:0008233; F: peptidase activity; IEA.
DR GO; GO:0005044; F: scavenger receptor activity; IEA.
DR GO; GO:0004295; F: trypsin activity; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR02172; LDL receptor A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PS00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Transmembrane; 1
FT NON_TER
SQ SEQUENCE 326 AA; 35561 MW; 8D6F12214393CDB2 CRC64;

Alignment Scores:
Pred. No.: 2,26e-48 Length: 326
Score: 649.00 Matches: 240
Percent Similarity: 50.13% Conservative: 50
Best Local Similarity: 36.94% Mismatches: 121
Query Match: 20.42% Indels: 68
DB: 13 Gaps: 8

US-09-323-597C-1 (1-1738) x Q7Z280 (1-326)
QY 439 GCGAGCAGTCTCAACTGGGATAGAGTGGACTCTCAGTACTGCATCAACCC 498
Db 1 GlyLeuArgSerCysSerGlyLysPheHisCysValSerValArgCysileSerArg 20
QY 499 TCTAATCGTGTGATGGGTCTCACATGCCCGCGGAGGACGAGATCGGTGTGT 558
Db 21 AsnAlaValCysAspGlyValGlnAspCysArgAspGlyGluAspGluLeuAsnCysVal 40
QY 559 CGCTCTACGACCAAACTCATCTCAGTGTACTCATCTCAGAGAAAGTCTCGCAC 618
Db 41 ArgValSerGlySerHisSerValLeuGlnValPheGly-----ArgGlyLeuTrpArg 58
QY 619 CCTGTGTCACAGACGACTGCAACGAGACTACGGCGGGCGGCTCGCGGACATGGC 678
Db 59 ThrValCysSerGluGlyTrpPASPSPSerGlnLeuSerThrLeuAlaCysA-rgGlnLeuGly 78
QY 679 TATAAGAATAATTTTACTCTAGCAGCAAGGAATAGTGGATGACAGCGGATCCACGAGTTT 738

```

```

Db 79 Tyr----- 79
QY 739 ATGAAGACTGAACACAGTGGCGCAATGTGCGATATCTATAAAACTGTACCACAGTGT 798
Db 79 ----- 79
QY 799 GCCTGTCTTCAAAAGCAGTGGTTCCTTTAGCGTGATAGCCTGCGGGGTCAACTTGAAC 858
Db 80 -----SerArgSerAlaIleSer----- 85
QY 859 TCAAGCGCGCAGACGACGATGTGGGGCGGAGAGCGCGCTCCCGGGGGCTGCGCCCTGG 918
Db 86 -----SerArgIleValGlyGlyAsnValSerIysSerGlyGlnValProTrp 101
QY 919 CAGTCAGCTGCAGCTCCAGAGCTCCAGTGTGCGGAGGCTCCATCATCATCCCGCGAG 978
Db 102 GlnValSerLeuHisTyrGlnAsnGlnTyrLeuCysGlyGlySerIleIleSerGluSer 121
QY 979 TGGATCTGACAGCGCCCACTGCTGGTGGAAAAACCTCTTAACAATCCATCGCATTCGACG 1038
Db 122 TrpIleLeuThrAlaAlaHisCysVal---PheGlyPheAlaGlnProValLeuTrpAsp 140
QY 1039 GCATTTTCGGGGATTTTACAGCAATCTTTTCATGTTCTTATGAGCGCGGATACCAAGTAGAA 1098
Db 141 ValTyrAlaGlyLeuIleAsnLeuPro---LeuSerLysAlaGluAlaHisSerValGlu 159
QY 1099 AAGTGTATTTCTCATCCAAATATGACTCCAGACCAAGAACATGATGATGATCGCTGATG 1158
Db 160 LysIleIleTyrHisAlaAsnPheArgSerLysSerPheSerTyrAspIleAlaLeuIle 179
QY 1159 AAGCTGCAAGAGCTCTGACTTTCAACGACTAGTGAACAGAGTGTCTGCGCCCAACCCA 1218
Db 180 LysLeuThrLeuProLeuThrPheAsnAspGluIleAlaProIleCysLeuProAsnTyr 199
QY 1219 GCGATGATGTCGACGACGACAGACTCTGCTGATTTCCGGTGGGGGGCGACGAGAG 1278
Db 200 GlyGluSerPheLysAsnGlyGlnMetCysLeuIleSerGlyTrpGlyAlaThrValAsp 219
QY 1279 AAGAGGAAGACCTCAGAGTGTGCAAGCTGCGCAAGTGTCTCTCATGAGACACAGAGA 1338
Db 220 SerGlyGluThrSerLeuSerLeuHisValAlaGlnValProLeuLeuSerAsnLysGlu 239
QY 1339 TGCAACAGCAGATATGTTCTATGACCACTGATCACACGACCGCATGATCTGCGCGGCTTC 1398
Db 240 CysArgLysLeuGly-----LeuThrAsnTrpAsnValCysThrGluPhe 254
QY 1399 CTGACGGGGAACGTGCGATTTCTCCAGGTGACAGTGGAGGCGCTCTGTCACCTCGAAG 1458
Db 255 LeuArgGly---ValGlyThrCysGlnGlyAspSerGlyGlyProLeu---AlaCysGln 272
QY 1459 AACAAATATCTGCTGCTGATAGGGGATACAAAGCTGGGGTTCTGCTGTGCCAAGCTTAC 1518
Db 273 GlySerAlaTrpThrLeuValGlyThrGlySerTrpAspGluAsnCysGlyLysValAsn 292
QY 1519 AGACGAGGAGTGTACGGGAATGTGATGTTTACCGGACTGGGATTTATCGACAAATG 1575
Db 293 LysProGlyIleTyrThrSerIleSerGluAlaLeuThrTrpIleGlnGluMet 311

RESULT 15
ID Q8CAN9 PRELIMINARY; PRT; 777 AA.
AC Q8CAN9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Protease (Fragment).
GN PSS7 OR A130097D21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 14:54:56 ; Search time 22 seconds
(without alignments)
8227.084 Million cell updates/sec

Title: US-09-323-597C-1
Perfect score: 3179
Sequence: 1 ggaggagcgaggcgagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model p/US09323597/runat 06072004 111727 29642/app query.fasta 1.1927
-Q=/cgn2 1/US09323597 -SUPFIX=rsfp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-DB=SwissProt 42 -OFMT=fastan -SUPFIX=blousum62 -TRANS=human40 cdi -LIST=45
-UNITS=bits -SPART=1 -END=-1 -WATRIX=blousum62 -ALIGN=15 -MODE=LOCAL
-DOALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09323597 @CNG 1 1 19 @runat 06072004 111727 29642 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2714	85.4	492	1 TMS2 HUMAN	Q15393 homo sapien
2	2117	66.6	490	1 TMS2 MOUSE	Q9jic8 mus musculus
3	880	27.7	453	1 TMS3 MOUSE	Q8k150 mus musculus
4	875.5	27.5	454	1 TMS3 HUMAN	P57727 homo sapien
5	680	21.4	435	1 TMS4 MOUSE	Q8vca5 mus musculus
6	676.5	21.3	437	1 TMS4 HUMAN	Q9nrs4 homo sapien
7	665.5	20.9	455	1 TMS5 MOUSE	Q9er04 mus musculus
8	660	20.8	1035	1 ENTK BOVIN	P98072 bos taurus
9	651.5	20.5	457	1 TMS5 HUMAN	Q9h343 homo sapien
10	647	20.4	1019	1 TMS5 HUMAN	P98073 homo sapien
11	638	20.1	1069	1 ENTK MOUSE	P97435 mus musculus
12	634	19.9	1034	1 ENTK PIG	P98074 sus scrofa
13	601	18.9	1042	1 CORI HUMAN	Q9Y5Q5 homo sapien
14	579	18.2	1113	1 CORI MOUSE	Q92319 mus musculus
15	575.5	18.1	855	1 ST14 HUMAN	Q9y5y6 homo sapien
16	569.5	17.9	638	1 XAL HUMAN	P03952 homo sapien
17	568.5	17.9	835	1 ST14 MOUSE	P56677 mus musculus
18	567	17.8	417	1 HEPS_HUMAN	P05981 homo sapien

19 556 17.5 638 1 KAL RAT
20 556 17.5 811 1 TMS6 MOUSE
21 552.5 17.4 812 1 PLAM_MOUSE
22 548.5 17.3 436 1 HEPS_MOUSE
23 538 16.9 638 1 KAL_MOUSE
24 536.5 16.9 811 1 TMS6 HUMAN
25 533 16.8 625 1 FALI1 HUMAN
26 531 16.7 416 1 HEPS RAT
27 527.5 16.6 418 1 HATT HUMAN
28 506 15.9 324 1 TEST_MOUSE
29 501 15.8 421 1 ACRO HUMAN
30 499.5 15.7 436 1 ACRO MOUSE
31 496 15.6 422 1 DESI HUMAN
32 490.5 15.4 415 1 ACRO PIG
33 486.5 15.3 761 1 NCTR_MOUSE
34 486.5 15.3 790 1 PLAM PIG
35 486 15.3 810 1 PLAM ERIEU
36 483.5 15.2 786 1 STUB DROME
37 479 15.1 311 1 TRYG MOUSE
38 475.5 15.0 431 1 ACRO RABIT
39 473.5 14.9 437 1 ACRO RAT
40 473.5 14.9 810 1 PLAM MACMU
41 471 14.8 290 1 PR27 HUMAN
42 467.5 14.7 321 1 TRYG HUMAN
43 465.5 14.6 306 1 BSS4_MOUSE
44 464.5 14.6 812 1 PLAM BOVIN
45 463 14.6 273 1 TRYT_SHEEP

ALIGNMENTS

RESULT 1
TMS2_HUMAN STANDARD; PRT; 492 AA.
ID TMS2_HUMAN
AC Q15393; Q9EXX1;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMSRSS2 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM M.A.
RX MEDLINE=97468144; PubMed=9325052;
RA Paoloni-Giacchino A., Chen H., Peitsch M.C., Rossier C.,
RA Antonarakis S.E.;
RT "Cloning of the TMSRSS2 gene, which encodes a novel serine protease
RT with transmembrane, LDLR, and SRCR domains and maps to 21q22.3.";
RL Genomics 44:309-320(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21309069; PubMed=11414763;
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
RT "Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364(2001).
RN [3]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=21139112; PubMed=11245484;
RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Safran D.C.,
RA Raitano A.B., Jakobovits A.;
RT "Catalytic cleavage of the androgen-regulated TMSRSS2 protease results
RT in its secretion by prostate and prostate cancer epithelia.";
RL Cancer Res. 61:1686-1692(2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;
RT "Expression of transmembrane serine protease TMSRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140(2001).

CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
 CC cleavage and secreted.
 CC -!- TISSUE SPECIFICITY: Expressed strongly in small intestine. Also
 CC expressed in prostate, colon, stomach and salivary gland.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -!- SIMILARITY: Contains 1 SRCR domain.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U75329; AAC51784.1; ..

CC EMBL: AF123453; AAD37117.1; ..

CC EMBL: AF270487; AAK29280.1; ..

CC HSSP: P00763; 1DPO.

CC MEROPS: S01.247; ..

CC Genew; HGNC:11876; TMPS22.

CC MIM; 602060; ..

CC GO; GO:0005887; C: integral to plasma membrane; TAS.

CC GO; GO:0008236; F: serine-type peptidase activity; TAS.

CC InterPro: IPR009003; Cys Ser trypsin.

CC InterPro: IPR002172; LDL receptor A.

CC InterPro: IPR001254; Peptidase_S1.

CC InterPro: IPR001314; Peptidase_S1A.

CC InterPro: IPR001190; Srcr_receptor.

CC Pfam; PF00089; trypsin; 1.

CC PRINTS; PR00722; CHYMOTRYPSIN.

CC SMART; SM00192; LDLA; 1.

CC SMART; SM00202; SR; 1.

CC SMART; SM00020; TRYPSIN; 1.

CC PROSITE; PS01209; LDLA_1; 1.

CC PROSITE; PS00068; LDLA_2; 1.

CC PROSITE; PS00420; SRCR_1; FALSE_NEG.

CC PROSITE; PS0287; SRCR_2; 1.

CC PROSITE; PS0240; TRYPSIN_DOM; 1.

CC PROSITE; PS00134; TRYPSIN_HIS; 1.

CC PROSITE; PS00135; TRYPSIN_SER; 1.

CC Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 CC Polymorphism.

CC CHAIN 1 255

CC TRANSMEMBRANE PROTEASE, SERINE 2, NON-
 CC CATALYTIC CHAIN.

CC TRANSMEMBRANE PROTEASE, SERINE 2,
 CC CATALYTIC CHAIN.

CC CYTOPLASMIC (POTENTIAL).

CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).

CC EXTRACELLULAR (POTENTIAL).

CC LDL-RECEPTOR CLASS A.

CC SRCR.

CC SERINE PROTEASE.

CC CHARGE RELAY SYSTEM (BY SIMILARITY).

CC CHARGE RELAY SYSTEM (BY SIMILARITY).

CC CHARGE RELAY SYSTEM.

CC CLEAVAGE (POTENTIAL).

CC BY SIMILARITY.

CC BY SIMILARITY.

CC BY SIMILARITY.

CC BY SIMILARITY.

CC BY SIMILARITY.

CC INTERCHAIN (BY SIMILARITY).

CC BY SIMILARITY.

CC BY SIMILARITY.

CC BY SIMILARITY.

CC N-LINKED (GLCNAC. . .) (POTENTIAL).

CC N-LINKED (GLCNAC. . .) (POTENTIAL).

CC K -> N (in dbSNP:1056602).

CC /FTId=VAR_011692.

CC R -> Q: LOSS OF CLEAVAGE.

CC 255

CC MUTAGEN

CC 255

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

FT MUTAGEN 441 441 S->A: LOSS OF ACTIVITY.
 FT CONFLICT 160 160 M -> V (IN REF. 3).
 FT CONFLICT 242 242 I -> L (IN REF. 1).
 FT CONFLICT 329 329 E -> Q (IN REF. 1).
 FT CONFLICT 489 489 RAD -> KAN (IN REF. 1).
 SQ SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;

Alignment Scores:

Pred. No.: 3,9e-187 Length: 492
 Score: 2714.00 Matches: 491
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.80% Mismatches: 0
 Query Match: 85.37% Indels: 0
 DB: 1 Gaps: 0

US-09-323-597C-1 (1-1738) x TMS2_HUMAN (1-492)

QY 112 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAAACACCGATGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAAACCCCTATCCCGCACGCCACCTGTGTGTCCTCCACTGTCTACGAGGTGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 QY 232 CCGGCTCAGTACTACCGCTCCCGGCGCCAGTACGCGCCGAGGTCCTGAGCAGCGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAACCCCGCTCTGCGACGCGCCCAATCCCATCCGGACAGTGTGCACCTCAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAGACACTGTGCATCAGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 GCGCTGGCTTACTCTCGAAGTTCATGGGCGAGCAAGTCTCCAACTCTGGGATAGAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCAGTACTGTGCATCAACCCCTCTAACTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGGGAGGACGAGAAATCGT 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
 QY 592 TACTCATCTCAGAGGAGTCTCGACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GCGGGGCGGCTGTGCGAGGACATGGGCTATAGAATAATTTTACTCTAGCCCAAGGAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATGCACGCGGATCCACAGCTTTATGAACCTTATGAACCTTATGAACCTTATGAACCTTAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAATACTGTACACAGT 831
 Db 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuLeuArg 240
 QY 832 TGTATAGCTGGGGGCTCAACTTGAACCTCAAGCCGCGAGCAGGAGTGTGGGCGCGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGCGCTCCGGGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCT 951
 Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280

```
QY 952 TGCGAGGCTCCATCATCACCCCGAGTGGATCGTGACGCGCCACTCGGTGGA AAA 1011
| | | | |
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
| | | | |
QY 1012 CCTCTTAACATCATCGATGACCGCGATTTCGGGGGTTTTGGAGCAATCTTTCATG 1071
| | | | |
Db 301 ProLeuAsnAsnProThrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
| | | | |
QY 1072 TTCTATCGAGCGGATACCAAGTAGAAAAGTGTCTCTCATCCAAATATGACTCCAAG 1131
| | | | |
Db 321 PheTyrglyAlaGlyTyrglnValGluLysValIleSerHisProIleThrAspSerLys 340
| | | | |
QY 1132 ACAAGAACATGATGCGTGTGATGAAGTGCAGAGGCTCTGACTTTCAACGACCTA 1191
| | | | |
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
| | | | |
QY 1192 GTGAACACAGTGTCTGCGCCCAACCCAGGATGATGTCGACGACAGACGCTCTGCTGG 1251
| | | | |
Db 361 ValIysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
| | | | |
QY 1252 ATTTCCGGTGGGGGGCCACGAGGAGAAAGGGAAGACCTCAGAGTGTGGAACGCTGCC 1311
| | | | |
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
| | | | |
QY 1312 AAGTGTCTCTCATGTGACACACAGATGACACAGAGATGATCTATGACACCTGATC 1371
| | | | |
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrglyValTyrglyAsnLeuIle 420
| | | | |
QY 1372 ACACGACCATGATCTGTGCGGCTCTCTGACGAGGAACTGCAATCTTCCAGGCTGAC 1431
| | | | |
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
| | | | |
QY 1432 AGTGAGGGCTCTGTGCTACTTGAAGACAAATATCTGTGCTGATGAGGGGTACAAGC 1491
| | | | |
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
| | | | |
QY 1492 TGGGTTCTCGCTGTGCAAGCTTACAGACACGAGGTGACGGAATGTGATGGTATTC 1551
| | | | |
Db 461 TrpGlySerGlyCysAlaLysAlaTyrglyArgProGlyValTyrglyAsnValMetValPhe 480
| | | | |
QY 1552 ACGACTGTGATTATTCGACAAATGAGGCGCAGACGC 1587
| | | | |
Db 481 ThrAspTrpIleTyrglyArgGlnMetArgAlaAspGly 492
| | | | |
RESULT 2
TMS2 MOUSE
ID TMS2_MOUSE STANDARD; PRT; 490 AA.
AC Q9J108; Q9JRC4; Q9QY82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
DE transmembrane protein X).
GN TMPSR52.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/c;
RX MEDLINE=20104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikho P.T.;
RT "Expression of transmembrane serine protease TMPSR2 in mouse and
RT human tissues."
RL J. Pathol. 193:134-140 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease x."
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=BALB/c;
RX MEDLINE=20104370; PubMed=10683448;
RA Jacquinet E.J., Rao N.V., Rao G.V., Hoidal J.R.;
RT "Cloning, genomic organization, chromosomal assignment and expression
RT of a novel mosaic serine proteinase: epitheliasin."
RL FEBS Lett. 468:93-100 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Mammary gland;
RX MEDLINE=22388457; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Willalson D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
CC cleavage and secreted (by similarity).
CC -1- TISSUE SPECIFICITY: Expressed mainly in prostate and kidney.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF199362; AAF97867.1; -
CC EXBL: AF243500; AAF64186.1; -
CC EXBL: AF113596; AAF21308.1; -
CC EXBL: BC038393; AAH38393.1; -
CC HSP: P00763; IDPO.
CC MEROPS: S01.247; -
CC MGD: MGI:1354381; Tmprss2.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: PF00057; ldl_recept_a; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00089; tryptsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLA; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00668; LDLRA_2; 1.
DR PROSITE: PS00420; SRCR_1; FALSE NEG.
DR PROSITE: PS00287; SRCR_2; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
KW CHAIN 1 253 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
```

FT CHAIN 254 490 CATALYTIC CHAIN.
 FT TRANSMEMBRANE PROTEASE, SERINE 2,
 FT CATALYTIC CHAIN.
 FT CYPOLASMIC (POTENTIAL).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT TRANSMEM 84 104 (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT LDL-RECEPTOR CLASS A.
 FT SRCR.
 FT DOMAIN 105 490 SERINE PROTEASE.
 FT 111 149 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT 150 242 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT 254 490 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT 294 294 CLEAVAGE (POTENTIAL).
 FT 343 343 BY SIMILARITY.
 FT 439 439 BY SIMILARITY.
 FT 253 254 BY SIMILARITY.
 FT 76 125 BY SIMILARITY.
 FT 119 138 BY SIMILARITY.
 FT 132 147 BY SIMILARITY.
 FT 171 230 BY SIMILARITY.
 FT 184 240 BY SIMILARITY.
 FT 243 363 INTERCHAIN (BY SIMILARITY).
 FT 279 295 BY SIMILARITY.
 FT 408 424 BY SIMILARITY.
 FT 435 463 BY SIMILARITY.
 FT 111 111 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT 212 212 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT 474 474 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 212 212 S -> L (IN REF. 3).
 FT 122 122 S -> N (IN REF. 3).
 FT 178 178 Y -> H (IN REF. 1).
 FT 320 320 Y -> D (IN REF. 1).
 FT 474 474 N -> D (IN REF. 1).
 FT CONFLICT 474 474
 FT CONFLICT 474 474
 SQ SEQUENCE 490 AA; 53479 MW; 07D2B03EA4D8A1A9 CRC64;

Alignment Scores:

Pred. No.: 2,648-144 Length: 490
 Score: 2117.00 Matches: 385
 Percent Similarity: 86.76% Conservative: 41
 Best Local Similarity: 78.41% Mismatches: 63
 Query Match: 66.59% Indels: 2
 DB: 1 Gaps: 2

US-09-323-597c-1 (1-1738) x TMS2 MOUSE (1-490)

QY 112 ATGGCTTTGAATCAGGCTCACCAGCTATGGACCTTACTATGAAACCATGGATAC 171
 Db 1 MetAlaLeuAenSer-GlySerProGlyIleGlyProCysTyrGluAsnHisGlyTyr 20
 QY 172 CAACGGGAACCCCTATCCGACAGCCACTGTGTCCTCCACTGTCTACAGGTGAT 231
 Db 21 GlnSerGluHisIleCysProArgProValAlaProAsnGlyTyrAsnLeuTyr 40
 QY 232 CGGCTCAGTACTACCTCCCTCCCTGCTCCAGTACGCGCGGCTCTCAGCGAGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgIleThrGlnAla 60
 QY 292 TCCAAACCCGCTGCTGACGAGCCAAATCCCATCCGAGACAGTGTGACCTCAAG 351
 Db 61 SerThrSerValIleHisThrHisProLysSer---SerGlyAlaProCysThrSerLys 79
 QY 352 ACTAAGAAAGCACTGTGATCCTTACCTTGGGACCTTCTCTGGGAGCTCGCGTG 411
 Db 80 SerLysLysSerLeuCysLeuAlaLeuAlaLeuGlyThrValLeuThrGlyAlaAlaVal 99
 QY 412 GCCTGCTGCTTCTGAAAGTTTCATGGGACAGAGTGTCTCCAACTCTGGGATAGTGC 471
 Db 100 AlaAlaValLeuLeuTyrArgPheTrpAspSerAenCysSerThrSerGluMetGluCys 119
 QY 472 GACTCTCAGTACTGATCAACCCCTTAACCTGATGTGATGGCTGTACACTGCCCC 531
 Db 120 GlySerSerGlyThrCysIleSerSerSerLeuTyrCysAspGlyValAlaHisCysPro 139
 QY 532 GCGGGGAGGACGAGAATCGTGTGCTCCCTCTACGACCAAACTTCTCTCAGTGC 591
 Db 140 AsnGlyGluAspGluAsnArgCysValArgLeuTyrGlyGlnSerPheIleLeuGlnVal 159

RESULT 3

TMS3 MOUSE

ID TMS3 MOUSE STANDARD; PRT; 453 AA.

AC Q8K1T0; Q8VDB0;

QY 592 TACTCATCTCAGAGAGTCTCTGGCACCCCTGTGTGCCAGAGCTGGAACAGAACTAC 651
 Db TyrSerGlnArgGlyAlaTrpTyrProValCysGlnAspAspTrpSerGluSerTyr 179
 QY 652 GCGGGGCGGCTCTCAGGAGCATGGGCTATAGGAATAATTTTACTTCTAGCAAGGAATA 711
 Db GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLeu 199
 QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACTGAACACAGTGGCGGCAATGTGAT 771
 Db ProAspGlnSerGlyAlaThrSerPheMetLysLeuAsnValSerSerGlyAsnValAsp 219
 QY 772 ATCTATAAAAACTGTACCAAGTATGCTCTTCAAAAGCAGTGGTTTCTTACGC 831
 Db LeuTyrLysLysLeuTyrHisSerAspSerCysSerSerArgMetValValSerLeuArg 239
 QY 832 TGTATAGCTCTCGGGGTCACTTGAACCTAAGCCGACAGACAGATTTGGGGCGGAG 891
 Db CysIleGluCysGlyValArg---SerValLysArgGlnSerArgIleValGlyLeu 258
 QY 892 AGCGCGCTCCCGGGGCTGGCCCTGGCAGCTCAGCTGCACGTCCAGAACGTCCACGTG 951
 Db AsnAlaSerProGlyAspTrpTrpGlnValSerLeuHisValGlnGlyValHisVal 278
 QY 952 TCGAGAGCTCCATCATCACCCCGAGTGTGACAGCGCCACTGCTGGAAAAA 1011
 Db CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluGlu 298
 QY 1012 CCTCTTAACAATCCATGCGATTCGACGCGATTTCGGGGATTTTGACAGATCTTTCATG 1071
 Db ProLeuSerGlyProArgTyrTrpThrAlaPheIleGlyIleLeuArgGlnSerLeuMet 318
 QY 1072 TTCTATGAGCCGGATACCAAGTAGAAAAAGTGTATTTCTCATCCAAATTTACTCCAAAG 1131
 Db PheTyrGlySerArgHisGlnValGlyLysValIleSerHisProAsnTyrAspSerLys 338
 QY 1132 ACCAAGAACATGACATTCGCTGATGAAGCTCGAAGCTCGAAGCTCTGACTTTCAGAGCTA 1191
 Db ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnThrProLeuAlaPheAsnAspLeu 358
 QY 1192 GTGAAACAGAGTGTCTGCCCAACCCAGGATGATGTGCGCCAGAACAGCTCTGTCTGG 1251
 Db ValLysProValCysLeuProAsnProGlyMetMetLeuAspLeuAspGlnGluCysTrp 378
 QY 1252 ATTTCGGTGGGGGCGCACCGAGGAGAAAGGAGACCTCAGAGTGTCTGAGCGTGGC 1311
 Db IleSerGlyTrpGlyAlaThrTyrGluLysGlyLysThrSerAspValLeuAsnAlaAla 398
 QY 1312 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACCAACTGATC 1371
 Db MetValProLeuIleGluProSerLysCysAsnSerLysTyrIleTyrAsnAsnLeuIle 418
 QY 1372 ACACAGCCATGATGTGTGCGGCTTCTCGAGGGAGAACCTCGATTTCTCCAGGGTGAC 1431
 Db ThrProAlaMetIleCysAlaGlyPheLeuGlnGlySerValAspSerCysGlnGlyAsp 438
 QY 1432 AGTGGAGGGCTCTGTGTCACCTTCAAGAACAAATATCTGGTGGCTGATAGGGGATACAAAG 1491
 Db SerGlyGlyProLeuValThrLeuLysAsnGlyIleTrpTrpLeuIleGlyAspThrSer 458
 QY 1492 TGGGGTCTGCTGTGCGCAAGCTTACAGACCGAGGTGTACGGGAATGTGATGGTATTC 1551
 Db TrpGlySerGlyCysAlaLysAlaLeuArgProGlyValTyrGlyAsnValThrValPhe 478
 QY 1552 ACCGACTGGATTTTACGACAAATCAGGCGCAGAC 1584
 Db ThrAspTrpIleTyrGlnGlnMetArgAlaAsn 489

FT	73	85	BY SIMILARITY.
DISULFID	73	85	BY SIMILARITY.
FT	79	98	BY SIMILARITY.
DISULFID	79	98	BY SIMILARITY.
FT	107	107	BY SIMILARITY.
DISULFID	107	107	BY SIMILARITY.
FT	129	194	BY SIMILARITY.
DISULFID	129	194	BY SIMILARITY.
FT	142	204	BY SIMILARITY.
DISULFID	142	204	BY SIMILARITY.
FT	207	324	BY SIMILARITY.
DISULFID	207	324	BY SIMILARITY.
FT	242	258	BY SIMILARITY.
DISULFID	242	258	BY SIMILARITY.
FT	338	406	BY SIMILARITY.
DISULFID	338	406	BY SIMILARITY.
FT	369	385	BY SIMILARITY.
DISULFID	369	385	BY SIMILARITY.
FT	424	424	BY SIMILARITY.
CARBOHYD	424	424	BY SIMILARITY.
FT	221	221	BY SIMILARITY.
CONFLICT	221	221	BY SIMILARITY.
FT	117	117	BY SIMILARITY.
CONFLICT	117	117	BY SIMILARITY.
FT	246	246	BY SIMILARITY.
CONFLICT	246	246	BY SIMILARITY.
FT	453	453	BY SIMILARITY.
SEQUENCE	453	453	BY SIMILARITY.

L -> H (IN REF. 1: CACB3350).
 V -> I (IN REF. 1: CACB3350).
 453 AA: 49491 MW; 1ABCBF10AF651EF6 CRC64;

Alignment Scores:
 Pred. No.: 1,45e-55 Length: 453
 Score: 880.00 Matches: 187
 Percent Similarity: 57.57% Conservative: 64
 Best local Similarity: 42.89% Mismatches: 159
 Query Match: 27.68% Indels: 26
 DB: 1 Gaps: 12

US-09-323-597C-1 (1-1738) x TMS3_MOUSE (1-453)

QY	316	CCCAATCCCATCGGCACAGTGTGCACCTCAAGACTAAGAAAGCACTGTGCATC---	372
QY <td>316</td> <td>CCCAATCCCATCGGCACAGTGTGCACCTCAAGACTAAGAAAGCACTGTGCATC---</td> <td>372</td>	316	CCCAATCCCATCGGCACAGTGTGCACCTCAAGACTAAGAAAGCACTGTGCATC---	372
DB	28	ProvalAlaProaspGlyAspAlaValAlaAlaGlnIleLeuSerLeuLeuProLeuLys 47	
QY <td>373</td> <td>-----ACCTTTGACCTCGGCACCTTCTCTGGGAGCTCGCTGGCGCGCTGGGCGCTA 422</td> <td></td>	373	-----ACCTTTGACCTCGGCACCTTCTCTGGGAGCTCGCTGGCGCGCTGGGCGCTA 422	
DB	48	PhePheProIleIleValIleGlyIleIleAlaLeuIleLeuAlaIleGlyLeu 67	
QY <td>424</td> <td>CTCTGGAGTTTCATCGGCAGCAAGTGTCTCCAACTCTGGGATAGTAGTCTCTCAGGT 483</td> <td></td>	424	CTCTGGAGTTTCATCGGCAGCAAGTGTCTCCAACTCTGGGATAGTAGTCTCTCAGGT 483	
DB	68	GlyIleHisPhe-----AspCysSerGlyLys---TyrArgCysHisSerSerPhe 83	
QY <td>484</td> <td>ACCTGTCATCAACCCCTCTTAACGTGTGTGATGGCGTGTACACTGCCCCGGCGGGAGAC 543</td> <td></td>	484	ACCTGTCATCAACCCCTCTTAACGTGTGTGATGGCGTGTACACTGCCCCGGCGGGAGAC 543	
DB	84	LysCysIleGluLeuThrAlaArgCysAspGlyValSerAspCysLysAsnAlaGluAsp 103	
QY <td>544</td> <td>GAGAACTCGGTGTGTGGCTCTACGGACCAAACTTCATCTCTCAGGTGTACTCATCTCAG 603</td> <td></td>	544	GAGAACTCGGTGTGTGGCTCTACGGACCAAACTTCATCTCTCAGGTGTACTCATCTCAG 603	
DB	104	GluTyrArgCysValArgValSerGlyGlnArgAlaAlaLeuGlnValPheThrAla--- 122	
QY <td>604</td> <td>AGGAAGTCTCTGGCACCCCTGTGTGCAAGACACTTGGAAACGAGAACTACGGCGCGCGGCC 663</td> <td></td>	604	AGGAAGTCTCTGGCACCCCTGTGTGCAAGACACTTGGAAACGAGAACTACGGCGCGCGGCC 663	
DB	123	---AlaAlaIlePheThrMetCysSerAspAspTrpLysSerHisTyrAlaLysIleAla 141	
QY <td>664</td> <td>TGCGGGGACATGGGCTATAAGAAATAATTTTACTCTAGCCAAAGAAATA----- 711</td> <td></td>	664	TGCGGGGACATGGGCTATAAGAAATAATTTTACTCTAGCCAAAGAAATA----- 711	
DB	142	CysAlaGlnLeuGlyPhePro---SerTyrValSerSerAspHisLeuArgValAspAla 160	
QY <td>712</td> <td>GTGATGACACGGGATCCACCAGCTTTATGAACTGAACACAAAGTGC CGGCAATGTCAT 771</td> <td></td>	712	GTGATGACACGGGATCCACCAGCTTTATGAACTGAACACAAAGTGC CGGCAATGTCAT 771	
DB	161	LeuGluGluGlnPheGlnGlyAspPheValSerIleAsnHisLeuLeuSerAsp---Asp 179	
QY <td>772</td> <td>ATCTATAAAAACCTGTACCACAGT-----GATGCCCTGTTCTTCANAAGCAGTG 819</td> <td></td>	772	ATCTATAAAAACCTGTACCACAGT-----GATGCCCTGTTCTTCANAAGCAGTG 819	
DB	180	LysValThrAlaLeuHisHisSerValTyrMetArgGluGlyCysThrSerGlyHisVal 199	
QY <td>820</td> <td>GTTTCTTACGCTGTATAGCTCGGGGTCAACTTGAAGTCAAGCGCCAGACGAGATT 879</td> <td></td>	820	GTTTCTTACGCTGTATAGCTCGGGGTCAACTTGAAGTCAAGCGCCAGACGAGATT 879	
DB	200	ValThrLeuLysCysSerAlaCysGlyThrArgThrGlyTyrSer-----ProArgIle 217	
QY <td>880</td> <td>GTGGCGCGCAGAGCGCGGCTCCCGGGGCGCTCGGCCCTGGCAGGTTCAGCTGCACGCTCAG 939</td> <td></td>	880	GTGGCGCGCAGAGCGCGGCTCCCGGGGCGCTCGGCCCTGGCAGGTTCAGCTGCACGCTCAG 939	
DB	218	ValGlyGlyAsnMetSerSerLeuThrGlnTrpProTrpGlnValSerLeuGlnPheGln 237	
QY <td>940</td> <td>AACGTCCACGTGTGCGGAGGCTTCATCATCACCCCGCAGGTGGATCGTGACAGCGGCCAC 999</td> <td></td>	940	AACGTCCACGTGTGCGGAGGCTTCATCATCACCCCGCAGGTGGATCGTGACAGCGGCCAC 999	
DB	238	GlyTyrHisLeuCysGlySerValIleThrProLeuThrIleValThrAlaIleHis 257	

QY 1000 TCGCTGCAAAACCTCTTACCAATCGATCGCATTCGGACGGCATTTGCGGGATTTTGAGA 1059
 Db CysVal---TyrAspLeuTyrHisProLysSerTrpThrValGlnValGlyLeuVal--- 275
 QY 1060 CAATCTTCATCTTCTATGAGGCC---GGATACCAAGTAGAAAAGTGAATTTCTCATCCA 1116
 Db ---SerLeuMetAspSerProValProSerHisLeuValGluLysIleIleTyrHisSer 294
 QY 1117 AATATGACTCCAGACCGAGCAATGACATTCGGCTGTGATGAGCTCAGAGCCTCTG 1176
 Db ---LysTyrLysProLysArgLeuGlyAsnAspIleAlaLeuMetLysLeuSerGluProLeu 314
 QY 1177 ACTTTCACGACCTAGTGAACCAAGTGTCTGCGCAACCCAGGCGATGCTGTCAGCCA 1236
 Db ThrPheAspGluThrIleGlnProIleCysLeuProAsnSerGluLysAsnPheProAsp 334
 QY 1237 GAACAGCTCTGCTGATTTCCGGTGGGGCCACCGAGGAGAAAGGAGACCTCAGAA 1296
 Db ---GlyLysLeuCysTrpThrSerGlyTrpGlyAlaThrGluAspGlyGlyAspAlaSerPro 354
 QY 1297 GTGCTGAACGCTGCCAAGGTCTCTCATTCGAGACACAGAGATGCAACAGCAGATATGTC 1356
 Db ValLeuAsnHisAlaAlaValProLeuLeuSerAsnLysIleCysAsnHisArgAspVal 374
 QY 1357 TATGCAACCTGATCACACCGACCATGATCTGTCCGGCTTCTCCAGGGAAGCTCGAT 1416
 Db TyrGlyGlyIleIleSerProSerMetLeuCysAlaGlyTyrLeuLysGlyGlyValAsp 394
 QY 1417 TCTTCCAGGTCAGAGTGGGGCTCTGCTGCTCCTCGAGAGCAATATCTGCTGGCTG 1476
 Db SerCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluArgLeuGluLysLeu 414
 QY 1477 ATAGGGGATACAAGCTGGGGTTCTGGCTGTCCAAAGCTTACAGACGAGGAGTGTACGG 1536
 Db ValGlyAlaThrSerPheGlyIleGlyCysAlaGluValAsnLysProGlyValTyrThr 434
 QY 1537 AATGTGATGTAATCACGAGCTGATTTATCGCAATATGAGGGGAGAC 1584
 Db ArgIleThrSerPheLeuAspTrpIleHisGluGlnLeuGluArgAsp 450

RESULT 4
 TMS3 HUMAN
 ID TMS3 HUMAN STANDARD; PRT; 454 AA.
 AC P57727;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
 TAG-12) (Tumor associated differentially-expressed gene-12 protein).
 GN TMPSR3 OR TAG12 OR ECHS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Sutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND T).
 RC TISSUE=Ovarian carcinoma;
 RX MEDLINE=20521358; PubMed=11068177;
 RA Underwood L.J., Sugimasa K., Tanimoto H., Beard J.B., Schneider B.N.,
 RA Wang Y., Farley T.H., O'Brien T.J.;
 RT "Ovarian tumor cells express a novel multi-domain cell surface serine
 protease";
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A, B AND D), AND VARIANT ILE-53.
 RX MEDLINE=20578749; PubMed=11137999;
 RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chrost R.,
 RA Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
 RA Younus F., Mehdi S.O., Radhakrishna U., Papasavas M.P., Gehrig C.,
 RA Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
 RA Antonarakis S.E.;
 RT "Insertion of beta-satellite repeats identifies a transmembrane
 RT protease causing both congenital and childhood onset autosomal

recessive deafness."; Nat. Genet. 27:59-63(2001).
 RN [3]
 RP SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
 RX MEDLINE=22281255; PubMed=12393794;
 RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
 RA Doughterty L., Vuagniaux G., Guida E., Okui M., Rossier C., Hancock M.,
 RA Buchet K., Raymond A., Hummel E., Marzella P.L., Kudoh J.,
 RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
 RT "The transmembrane serine protease (TMPSR3) mutated in deafness
 RT DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro."; Hum. Mol. Genet. 11:2829-2836(2002).
 RN [4]
 RP VARIANTS DFNB8/DFNB10 CYS-251 AND LEU-404.
 RX MEDLINE=21354467; PubMed=11462234;
 RA Masoudi S., Antonarakis S.E., Schwede T., Ghorbel A.M., Gratri M.,
 RA Papasavas M.P., Driha M., Eigaied-Bouilla A., Wattenhofer M.,
 RA Rossier C., Scott H.S., Ayadi H., Guipponi M.;
 RT "Novel missense mutations of TMPSR3 in two consanguineous Tunisian
 RT families with non-syndromic autosomal recessive deafness."; Hum. Mutat. 18:101-108(2001).
 RN [5]
 RP VARIANTS DFNB8/DFNB10 TRP-109, PHE-194 AND ARG-407, AND VARIANTS
 RP ILE-53; SER-111 AND VAL-253.
 RX MEDLINE=21317610; PubMed=11424922;
 RA Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,
 RA Kudoh J., Shibuya K., Antonarakis S.E., Bonne-Tamir B.,
 RA Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,
 RA Nance W.S., Wilcox R.R., Friedman T.B., Morell R.J.;
 RT "Novel mutations of TMPSR3 in four DFNB8/B10 families segregating
 RT congenital autosomal recessive deafness."; J. Med. Genet. 38:396-400(2001).
 RN [6]
 RP VARIANTS DFNB8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.
 RX MEDLINE=21904597; PubMed=11907649;
 RA Wattenhofer M., Di Iorio V., Rabinet R., Dougherty L., Panpanos A.,
 RA Schwede T., Montserrat-Sentis B., Arbones L., Iliadis T.,
 RA Pasquabisciglio A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M.,
 RA Petersen M.B., Estivill X., Gasparini P., Scott H.S.,
 RA Antonarakis S.E.;
 RT "Mutations in the TMPSR3 gene are a rare cause of childhood
 RT nonsyndromic deafness in Caucasian patients."; J. Mol. Med. 80:124-131(2002).
 RL J. Mol. Med. 80:124-131(2002).
 CC -I- FUNCTION: Probable protease. Seems to be capable of activating
 CC ENaC.
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
 CC reticulum.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=A;
 CC IsoId=P57727-1; Sequence=Displayed;
 CC Name=B; Synonyms=C;
 CC IsoId=P57727-2; Sequence=VSP_005391;
 CC Name=D;
 CC IsoId=P57727-3; Sequence=VSP_005392;
 CC Name=T; Synonyms=Truncated; TAG-12V;
 CC IsoId=P57727-4; Sequence=VSP_005393; VSP_005394;
 CC -I- TISSUE SPECIFICITY: Expressed in many tissues including fetal
 CC cochlea. Isoform T is found at increased levels in some
 CC carcinomas.
 CC -I- PTM: Undergoes autophosphorylation.
 CC -I- DISEASE: Defects in TMPSR3 are a cause of childhood-onset
 CC autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].
 CC -I- DISEASE: Defects in TMPSR3 are a cause of congenital autosomal
 CC recessive neurosensory deafness 10 (DFNB10) [MIM:605316].
 CC -I- SIMILARITY: Belongs to peptidase family S1.
 CC -I- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -I- SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

QY 955 GGAGGCTCATCATCACCCCGAGTGGATGCTGAGCGCGCCACCTGCGTGGAAAAACCT 1014
 Db 243 GlyGlySerValIleThrProLeuTrpIleIleThrAlaAlaHisCysVal---TyrAsp 261
 QY 1015 CTTAAACAATCCATGCGCATTTGGACGCGCAATTTGGACGCAATCTTTCATGTC 1074
 Db 262 LeuTyrLeuProLeuSerTrpThrIleGlnValGlyLeuVal-----SerLeuLeuAsp 279
 QY 1075 TATGGAGCC---GGATACCAAGTAGAAAAGTATTTCTTCAATCAAAATGATCCCAAG 1131
 Db 280 AsnProAlaProSerHisLeuValGluValIleValTyrHisSerLysTyrLysProLys 299
 QY 1132 ACCAAGAACATGATGCGTGTGATGAGTGCAGAGCTCTCTGACTTTCACAGCACTA 1191
 Db 300 ArgLeuGlyAsnAspIleAlaLeuMetLysLeuAlaGlyProLeuIleThrPheAsnGluMet 319
 QY 1192 GTGAACACAGTGTCTGCCAACCCAGGCGATGATGTCGACGCGACACACTCTGCTGG 1251
 Db 320 IleGlnProValCysLeuProAsnSerGluGluAsnPheProAspGlyLysValCysTrp 339
 QY 1252 ATTTCCGGTGGGGCCACCGAGGAG---AAAGGAGAGACTCAGAGTCTGCAAGCT 1308
 Db 340 ThrSerGlyTrpGlyAlaThrLeuAspGlyAlaGlyAspAlaSerProValLeuAsnHis 359
 QY 1309 GCCAAGTGTCTTCTCATGACACACAGATGCCACGATATGCTATGACAACTG 1368
 Db 360 AlaAlaValProLeuIleSerAsnLysIleCysAsnHisAspValTyrGlyIle 379
 QY 1369 ATCACACAGCATGATGTCGGCTTCTGCGAGGAGACGTCGATCTTTCAGGCT 1428
 Db 380 IleSerProSerMetLeuCysAlaGlyTyrLeuThrGlyGlyValAspSerCysGlnGly 399
 QY 1429 CACAGTGGAGGCGCTCTGCTCACTTCGAAGAACATATCTGCTGCTGATAGGGGATACA 1488
 Db 400 AspSerGlyGlyProLeuValCysGlnGluArgIleTyrLysLeuValGlyAlaThr 419
 QY 1489 ACCTGGGGTCTGCTGTCGCAAGCTTACAGACAGAGTGTACGGGAATGTATGATGTA 1548
 Db 420 SerPheGlyIleGlyCysAlaGluValAsnLysProGlyValTyrThrArgValThrSer 439
 QY 1549 TTCACGGAAGTGGATTATTCGCAATGAGGCGGACAG 1584
 Db 440 PheLeuAspTrpIleHisGluGlnMetGluArgAsp 451
 RESULT 5
 TMS4 MOUSE
 ID TMS4 MOUSE
 AC Q8VCA5;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Channel-activating
 DE protease 2) (mCAP2).
 GN TMPSR4 OR CAP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22144321; PubMed=12149280;
 RA Vagniaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
 RT "Synergistic activation of ENaC by three membrane-bound channel-
 RT activating serine proteases (mCAP1, mCAP2, and mCAP3) and serum- and
 RT glucocorticoid-regulated kinase (SGK1) in Xenopus oocytes.";
 RL J. Gen. Physiol. 120:191-201(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22389257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gurnarane P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Rodriguez R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Probable protease. Seems to be capable of activating
 CC ENaC.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AY043240; AAK85307.1; -;
 DR EMBL; BC021368; A3H21368.1; -;
 DR HSSP; P00761; IAN1.
 DR MEROPS; S01.034; -;
 DR MGD; MGI:2384877; Tmprss4.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR0012172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1-
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; P00722; CHYMOTRYPSIN.
 DR PROSITE; PS01209; LDLRA_1; FALSE_NEG.
 DR PROSITE; PS00668; LDLRA_2; FALSE_NEG.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 FT DOMAIN 1 30 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 31 51 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 52 435 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 59 101 LDL-RECEPTOR CLASS A.
 FT DOMAIN 102 202 SRCR.
 FT DOMAIN 203 435 SERINE PROTEASE.
 FT ACT_SITE 243 243 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 268 288 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 385 395 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 202 203 CLEAVAGE (POTENTIAL).
 FT DISULFID 62 81 BY SIMILARITY.
 FT DISULFID 75 90 BY SIMILARITY.
 FT DISULFID 125 181 BY SIMILARITY.
 FT DISULFID 138 191 BY SIMILARITY.
 FT DISULFID 194 308 BY SIMILARITY.
 FT DISULFID 228 244 BY SIMILARITY.
 FT DISULFID 354 370 BY SIMILARITY.
 FT DISULFID 381 408 BY SIMILARITY.

FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 435 AA; 47495 MW; DC52B45A43E01369 CRC64;
Alignment Scores:
Pred. No.: 3.21e-41 Length: 435
Score: 680.00 Matches: 163
Percent Similarity: 50.56% Conservative: 61
Best Local Similarity: 36.79% Mismatches: 173
Query Match: 21.39% Indels: 46
DB: 1 Gaps: 13
US-09-323-597C-1 (1-1738) x TMS4_MOUSE (1-435)
QY 313 CAGCCCAATCCCATCGGACAGTGTGCACCTCAAGACTAGAAGCA----- 363
Db LysProArgPheGlnGluThrPhe-----LysLysValGlyIlePro 33
QY 364 ---CTGGCATCACCTGACCTCGGACCTTCTCGTGGGAGCTCGCGCGCGCGCGG 420
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 421 CTACTCGGAATTCATCGGACAGCAAGTGTCTCAACTCTGGGATAGAGTGGACTCTCA 480
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 481 GGTACTGTCATCAACCCCTCAACTGTTGATGGCGGTGTGCACACTGCCCGCGGCGG 540
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 541 GACGAGAAATCGGTGT-----GTTCGCTCTAC 567
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 568 GGACCAAACTCATCTCAGTGTACTCATCTCAGAGGAGTCTCGCACCTGTGTC 627
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 628 CAAGACGACTGGAACGAGAACTACGCGCGCGCGCGCTGCGAGGACATGGGTATAA 687
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 688 AAT---TTTACTCTAGCAGGAAATAGTGGATGACAGCGGATCCACCGCTTTATG 744
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 745 CTGAACACAACTGCGCGCAATGTCATCTATAAAACCTGACACAGTGTGTC 804
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 805 TCTTCAAAAGCAGTGTCTTCTTACGTTGATAGCTGCGCGGTCACCTTGAATCAAG 864
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 865 CGCCAGACGAGATGTGGCGCGGAGAGCGCGCTCCCGCGCGCTGCGCGAGTC 924
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 925 AGCTGACGCTCCAGAGCTCCAGTGTGGAGGCTCCATCATCCCGAGTGGATC 984
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 985 GTGACAGCGCGCTCGTGGAAACCTTAACTTAACTTAACTTAACTTAACTTAACT 1041
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 1042 TTTGCGGGGATTTGAGCAATCTTTCATGTTCTATGAGCGCGGATACCAAGTAGAAA 1101
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 1102 GTG---ATTCTCATTCGAATATGACTCCAGACCAAGCAATGATGATGATGATG 1158
Db LysLysValGlyIlePro-----LysLysValGlyIlePro 33
QY 1274 ILEPHEILLEALAGLUProAsn---ProLeuTyrProLysGluLysAspIleAlaLeuVal 292

RESULT 6

TMS4_HUMAN
ID TMS4_HUMAN STANDARD; PRT; 437 AA.
AC Q9NR84; Q9NZAS;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
protease 2) (WP-SP2).
GN TMPSR4 OR TMPSR3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwamura T., Ruthenburger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMPSR3) overexpressed in
pancreatic cancer";
RL Cancer Res. 60:2602-2606(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Snekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
trachea, colon, and small intestine: identification, cloning, and
chromosomal localization";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Murzyn D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton B., Kettunen M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalilus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: Probable protease. Seems to be capable of activating ENAC (By similarity).

CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).

CC -!- TISSUE SPECIFICITY: High levels in pancreatic, gastric, colorectal and ampullary cancer. Very weak expression in normal gastrointestinal and urogenital tract.

CC -!- SIMILARITY: Belongs to peptidase family S1.

CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.

CC -!- SIMILARITY: Contains 1 SRCR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF179224; AAF74526.1; -

EMBL; AF216312; AAF31436.1; -

EMBL; BC011703; AAF11703.1; -

HSSP; P00763; 1DPO.

DR MEROPS; S01.034; -

DR GENE; HGNC:11878; TMPSRS4.

DR MIM; 606565; -

DR GO; GO:0016021; C:integral to membrane; NAS.

DR GO; GO:0004252; F:serine-type endopeptidase activity; NAS.

DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR002172; LDL_receptor_A.

DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase S1A.

DR InterPro; IPR001190; Srcr receptor.

DR Pfam; PF00057; ldl_recept_a; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00182; LDLr; 1.

DR SMART; SM00202; SR; 1.

DR SMART; SM00020; Tryp_SPC; 1.

DR PROSITE; PS01209; LDLR_1; FALSE NEG.

DR PROSITE; PS00068; LDLR_2; FALSE NEG.

DR PROSITE; PS00420; SRCR_1; FALSE NEG.

DR PROSITE; PS02287; SRCR_2; 1.

DR PROSITE; PS02240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.

DOMAIN 1 32

FT DOMAIN 33 53

FT TRANSMEM 33 53

FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

FT DOMAIN 54 437

FT DOMAIN 61 93

FT DOMAIN 94 204

FT LDL-RECEPTOR CLASS A. SRCR.

FT DOMAIN 205 437

FT ACT SITE 245 245

FT ACT SITE 290 290

FT ACT SITE 387 387

FT ACT SITE 204 205

FT CLEAVAGE (POTENTIAL).

FT SITE 64 83

BY SIMILARITY.

FT DISULFID 77 92 BY SIMILARITY.

FT DISULFID 127 183 BY SIMILARITY.

FT DISULFID 140 193 BY SIMILARITY.

FT DISULFID 196 310 BY SIMILARITY.

FT DISULFID 230 246 BY SIMILARITY.

FT DISULFID 356 372 BY SIMILARITY.

FT DISULFID 383 410 BY SIMILARITY.

FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 1 31 MLOPDSQPLNSLDVPLRKPRIPEMTFRK -> MGNPCA NPVSPMRPSS (IN REF. 2).

SQ SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;

Alignment Scores:

Pred. No.: 5,72e-41 Length: 437

Score: 676.50 Matches: 150

Percent Similarity: 53.91% Conservative: 57

Best Local Similarity: 39.06% Mismatches: 128

Query Match: 21.28% Indels: 49

DB: 1 Gaps: 13

US-09-323-597C-1 (1-1738) x TMS4_HUMAN (1-437)

Qy 508 TGTGATCGCTGTACACTCCCGCGCGGAGSACGAGAAATCGGTGTCGCTCTAC 567

Db 77 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysVallySerPhe 96

Qy 568 -----CGACCA----- -AATTTCATCTTCAGTGTAC 594

Db 97 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuInValLeu 116

Qy 595 TCATCTCAGAGGAAGTCTCGCACCCCTGTGTGCCAAGACGACTGGAAACGAACTACGGG 654

Db 117 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 136

Qy 655 CGGCGCGCTCGAGGACATGGCTAT-----AAGATATAATTTTACTCTAGCCAA--- 705

Db 137 GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluLe 156

Qy 706 -----GGAATAGTGGATGACAGCGGATCCACAGCTTATGAACATG 747

Db 157 GlyProAspGlnAspLeuAspValValGluLeuThrGluAsnSerGlnGluLeuArgMet 176

Qy 748 AACACAGTCCCGGCATGTGATATCTATAAAAACCTGTACCACAGTGTGCTGTCT 807

Db 177 ArgAsnSerSergly----- -ProCysLeu 184

Qy 808 TCAAAAGCAGTGTCTTTTACGCTGTATAGCTCGGGGTCAACTTGAACCTCAAGCCGC 867

Db 185 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 201

Qy 868 CAGACAGGATTGTGGCGCGGAGAGCGCTCCCGGGGGCCTGGCCCTGGCAGGTGAGC 927

Db 202 ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer 221

Qy 928 CTCACGTCAGAACGTCACAGTGTGGAGGCTCCATCATCATCACCCTGGCAGTGTGCTG 987

Db 222 IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 241

Qy 988 ACAGCCGCCACTGCGTGGGAAAACCTCTTAAACAATTCATGGCATTTGGACGGCATTTGCG 1047

Db 242 ThrAlaAlaHisCysPheArgLysHis--ThrAspValPheAsnTrpLysValArgAla 260

Qy 1048 GGGATTTCAGAA---CAATCTTTTCATGTTCTATGGAGCGCGGATACCAAGTAGAAAAGTG 1104

Db 261 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 276

Qy 1105 ATT-----TCTCATCAAAATTATGACTCCACAGACCAAGAACCAATGACATTGGCGCTG 1155

Db 277 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 293

Qy 1156 ATGAAGTCGAGAAAGCTCTGACTTTCAACGACCTAGTGAACCAAGTGTGTGCTGCCAAC 1215

Db 294 MetLysIeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 313
Qy 1216 CCAGGCATGATGCTGCACGACAGAGCTCTGCTGGATTTCGGGGGGCCACCGAG 1275
Db 314 PheAspGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 333
Qy 1276 GAGAAA---GGAGAGCTCAGAGTCTGACCTGCACAGGCTTCATTCATGACAA 1332
Db 334 GlnAsnGlyGlyLysMetSerAspIleLeuGlnValGlnValIleAspSer 353
Qy 1333 CAGAGATGCAACAGCATATGCTATGACACACCTGATCACACAGCATGATCTGTGCC 1392
Db 354 ThrArgCysAsnAlaAspAlaThrGlnGlyGluValThrGluLysMetCysAla 373
Qy 1393 GGCTTCCTCAGAGGAGAGCTGATCTGTCAGGAGTGCAGAGTGCAGAGGCTTCGTCACT 1452
Db 374 GylleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetYr 393
Qy 1453 TCGAAGAACAAATATCTGCTGCTCATAGGGATACAGCTGGGTCTGGTGGCCAAA 1512
Db 394 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyGlyCysGlyGly 412
Qy 1513 GCTTACAGACAGAGTGTGCGGAATGTGATGATTCACGAGTGTGATTCGACAA 1572
Db 413 ProSerThrProGlyValThrLysValSerAlaThrLeuAsnTrpIleTyrAsnVal 432
Qy 1573 ATGAGGGCAGAC 1584
Db 433 TrpLysAlaGlu 436

RESULT 7
TMS5 MOUSE
ID TMS5 MOUSE STANDARD; PRT; 455 AA.
AC Q9ER04; Q9ER02; Q9ER03;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
GN TMSR95.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC Mitsui S., Yamaguchi N.;
RT "cDNA cloning of mouse spinesin";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RA Mitsui S., Yamaguchi N.;
RT "Molecular cloning of mouse type 4 spinesin";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=4;
CC IsoId=Q9ER04-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9ER04-2; Sequence=VSP_005397, VSP_005398;
CC Name=2;
CC IsoId=Q9ER04-3; Sequence=VSP_005395;
CC Name=3;
CC IsoId=Q9ER04-4; Sequence=VSP_005396;
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; AB016229; BAB20276.1; -
DR EMBL; AB016230; BAB20277.1; -
DR EMBL; AB016423; BAB20278.1; -
DR EMBL; AB041037; BAB40328.1; -
DR HSP; P00763; IDPO.
DR MEROPS; S01.313; -
DR MGD; MGI:1933407; TmpRSS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS02087; SRCR_2; 1.
DR Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 71 455 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 112 207 SRCR.
FT DOMAIN 218 455 SERINE PROTEASE.
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 218 CLEAVAGE (POTENTIAL).
FT DISULFID 135 196 BY SIMILARITY.
FT DISULFID 148 206 BY SIMILARITY.
FT DISULFID 209 328 BY SIMILARITY.
FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT DISULFID 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPIC 1 144 Missing (in isoform 2).
FT VARSPIC 1 10 Missing (in isoform 3).
FT VARSPIC 1 182 Missing (in isoform 1).
FT VARSPIC 183 192 Missing (in isoform 1).
FT CONFLICT 325 325 D -> G (IN REF. 1; BAB20277).
FT SEQUENCE 455 AA; 49632 MW; 5CFC31789C6899AA CRC64;
SQ
Alignment Scores:
Pred. No.: 3,55e-40 Length: 455
Score: 665.50 Matches: 145
Percent Similarity: 50.97% Conservative: 66
Best Local Similarity: 35.02% Mismatches: 176
Query Match: 20.93% Indels: 27
DB: 1 Gaps: 9
US-09-323-597C-1 (1-1738) x TMS5_MOUSE (1-455)
Qy 367 TGCATCACCTTGCACCTCGGGACCTTCCTCGTGGAGCTGCGTGGCCGCTGCTACTC 426
Db 49 CysValIleLeuGlyValLeuGlyLeuLeuAlaGlyAlaGlyIleAlaSerTrpLeuLeu 58
Qy 427 TGGAGTTCATCGGAGCAGCAGTGTCCAACTCTGGATAGATGGAGCTCTCAGGTACC 486
Db 69 ValLeuTyrLeuTrpProAlaAlaSerProSer-----IleSerGlyThr 83

```
QY 487 TGCATCAACCCCTCTAACTGGTGTGATGGGTGTGCACACTGCCCGCGGGAGACAG 546
Db 84 LeuGlnGluGlu-----GluMetThrLeuAsnCySerProGlyValSerCysGlu 99
QY 547 AATCGGTGTGTT-----CGCCTCTACGCAACCAATTC 579
Db 100 GluGluLeuLeuProSerLeuProLysThrValSerPheArgIleAsnGlyGluAspLeu 119
QY 580 ATCCCTTCAGGTACTCATCTCAGAGGAGTCTCGCACCTGTGTGTCAGACACTGG 639
Db 120 LeuLeuGlnValGlnValAlaArgProAspTrpLeuLeuValCysHisGluGlyTrp 139
QY 640 AACGAGAACTACGGCGCGCGCTGCAGGACATGGGCTAT---AAGAATAATTTTAC 696
Db 140 SerProAlaLeuGlyMethIleCysLysSerLeuGlyHisIleArgLeuThrGlnHis 159
QY 697 TCTAGCCAAAGGATAGTGAATGAC-----ACCGATCCACGACCTTTATGAACTGAAC 750
Db 160 LysAlaValAsnLeuSerAspIleLysLeuAsnArgSerGlnGluPheAlaGlnLeuSer 179
QY 751 ACAAGTCCGCGCAATGTGATATCTATAAAAGTGTACCAAGCATGATCCCTGTCTTCA 810
Db 180 AlaArgProGlyGlyLeu---ValGluGluAlaTrpLysProSerAlaAsnCySerProSer 198
QY 811 AAAGCAGTGGTTCTTTACGTGTATAGCTGTATAGCTCGGGGTCAACTGGAATCAAGCCGCCAG 870
Db 199 GlyArgIleValSerLeuLysCysSerGlyCysGlyAlaArg-----ProLeuAla 215
QY 871 ACAGCATTTGCGGGCGGAGCGCGCTCCCGGGCGCTGCCCTGCGCAGCTCAGCTG 930
Db 216 SerArgIleValGlyGlyGlnAlaValAlaSerGlyArgTrpProTrpGlnAlaSerVal 235
QY 931 CACGTCACAGCTCCATCGTGTGCGAGGCTCCATCATCACCCCGATGGATCGTGACA 990
Db 236 MetLeuGlySerArgHisThrCysGlyAlaSerValLeuAlaProHisTrpValValThr 255
QY 991 GCGGCGCATCGGTGGAAGAAA---CCTCTTACAATCCATGCGCATTTGGACGCCATTGCG 1047
Db 256 AlaAlaHisCysMetTyrSerPheArgLeuSerArgLeuSerSerTrpArgValHisAla 275
QY 1048 GGGATTTTGAGACAATCTTTTCATGTCTATGAGCGCGGATACCAAGTAGAAGAAAGTATT 1107
Db 276 GlyLeuValSerHisGlyAlaValArgGlnHisGlnGlyThrMetValGluLysIleIle 295
QY 1108 TCTCATCAAAATATGACTCCAGACCAAGACAATGATGATTCGCTGATGAGCTGACG 1167
Db 296 ProHisProLeuTyrSerAlaGlnAsnHisAspTyrAspValAlaLeuLeuGlnLeuArg 315
QY 1168 AAGCCTCTGACTTTCAACGACCTAGTGAACACAGTGTGTCTCCCAACCCAGCGCATGATG 1227
Db 316 ThrProIleAsnPheSerAspThrValAspAlaValCysLeuProAlaLysGluGlnTyr 335
QY 1228 CTGCGACCCAGAACGCTCTGCTGATTTCCGGGTGGGGGCCACCCGAG---GAGAAAGGG 1284
Db 336 PheProTrpGlySerGlnCysTrpValSerGlyTrpGlyHisThrAspProSerHisThr 355
QY 1285 AAGACTCAGAGTGTGAACGCTCCAGGTGCTTCTCATTTGACACAGACAGATGCAAC 1344
Db 356 HisSerSerAspThrLeuGlnAspThrMetValProLeuLeuSerThrHisLeuCysAsn 375
QY 1345 ACCGATATGCTATGACACCTGATCACACCGACCATGATCTGTGCGCGTCTCTGCGAG 1404
Db 376 SerSerCysMetTyrSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyTyrLeuAsp 395
QY 1405 GGGACGTCGTCTCTGCGAGGTGACAGTGGAGGCTCTGCTCACTTCGAGAACAT 1464
Db 396 GlyArgAlaAspAlaCysGlnGlyAspSerGlyGlyProLeuValCysProSerGlyAsp 415
QY 1465 ATCTGGTGGCTATAGGGGATACAAAGCTGGGGTCTGCTGTGCCAAAGCTTACAGACCA 1524
Db 416 ThrTrpHisLeuValGlyValValSerTrpGlyArgGlyCysAlaGluProAsnArgPro 435
```

1525 GGAGTGTACGGGAATGTGATGGTATTCACGCACTGGATTTAT 1566
436 GlyValTyrAlaLysValAlaGluPheLeuAspTrpIleHis 449

RESULT 8
ENTK_BOVIN
ID ENTK_BOVIN STANDARD; PRT: 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duoenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu O., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043122; PubMed=8226855;
RA Lavallie B.R., Rehemtulla A., Racie L.A., Diblasio E.A.,
RA Perenz C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
subunit of bovine enterokinase.";
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
enterokinase.";
RL J. Protein Chem. 10:475-480(1991).
CC -!- FUNCTION: Responsible for initiating activation of pancreatic
proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
A). It catalyzes the conversion of trypsinogen to trypsin which in
turn activates other proenzymes including chymotrypsinogen,
procarboxypeptidases, and proelastases.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
trypsinogen.
CC -!- SUBUNIT: Heterodimer of a catalytic (light) chain and a
multidomain (heavy) chain linked by a disulfide bond.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P98072-1; Sequence=displayed;
CC Name=Short;
CC IsoId=P98072-2; Sequence=VSP_005386;
CC -!- TISSUE SPECIFICITY: Intestinal brush border.
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPsin-LIKE PROTEASE.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 1 SSCR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
CC ENBL; U09859; AAB40026.1; -
CC ENBL; L19663; AAA16035.1; -
CC PIR; A43090; A43090.
CC PDB; 1EKB; 14-OCT-99.
CC MEROPS; S01.156; -.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR000998; MAM domain.
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR001314; Peptidase S1A.
CC InterPro; IPR000082; SEA domain.
CC InterPro; IPR001190; Src_receptor.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00057; ldl_recept_a; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF01390; SEA; 1.
CC Pfam; PF00530; SRCR; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00261; LDLRECEPTOR.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00192; LDLa; 2.
CC SMART; SM00137; MAM; 1.
CC SMART; SM00200; SEA; 1.
CC SMART; SM00202; SR; 1.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS01180; CUB; 2.
CC PROSITE; PS01209; LDLRA_1; 2.
CC PROSITE; PS00668; LDLRA_2; 2.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS00660; MAM_2; 1.
CC PROSITE; PS00024; SEA_1.
CC PROSITE; PS00420; SRCR_1; FALSE_NEG.
CC PROSITE; PS00287; SRCR_2; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Signal-anchor; Glycoprotein; Myristate; Hydrolase; Serine protease;
CC Zymogen; Transmembrane; Repeat; Alternative splicing; 3D-structure;
CC Lipoprotein.
CC CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
CC CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
CC DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
CC DOMAIN 48 1035 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 54 169 SEA.
CC DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
CC DOMAIN 240 350 CUB 1.
CC DOMAIN 358 520 MAM.
CC DOMAIN 540 650 CUB 2.
CC DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
CC DOMAIN 694 787 SRCR.
CC DOMAIN 801 1035 SERINE PROTEASE.
CC ACT_SITE 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 987 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC LIPID 2 2 N-myristoyl glycine (Potential).
CC DISULFID 199 212 BY SIMILARITY.
CC DISULFID 205 225 BY SIMILARITY.
CC DISULFID 219 236 BY SIMILARITY.
CC DISULFID 659 671 BY SIMILARITY.
CC DISULFID 666 684 BY SIMILARITY.
CC DISULFID 678 693 BY SIMILARITY.
CC DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
CC DISULFID 826 842 BY SIMILARITY.
CC DISULFID 926 993 BY SIMILARITY.
CC DISULFID 957 972 BY SIMILARITY.
CC DISULFID 983 1011 BY SIMILARITY.

```

```

FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 486 486 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 864 864 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARGPLIC 166 192 Missing (in isoform Short).
FT CONFLICT 808 808 R -> Y (IN REF 3).
SQ SEQUENCE 1035 AA; 114887 MW; 5207970B08296E13 CRC64;

Alignment Scores:
Pred. No.: 1.05e-39 Length: 1035
Score: 660.00 Matches: 150
Percent Similarity: 48.98% Conservative: 67
Best Local Similarity: 33.85% Mismatches: 180
Query Match: 20.78% Indels: 46
DB: 1 Gaps: 12

US-09-323-597C-1 (1-1738) x ENT_K_BOVIN (1-1035)
QY 304 GTCGTCAGCAGCCCAATCCCATCCGGACAGTGCACCTCAAGACTAAGAAGCA 363
DB 611 ValTyrThrGlyProGlyProValAsnAspValPheSerThrThrAsnArgMetThrVal 630
QY 364 CTGTGCATCACCTTGACCTG-----GGACCTTCCTCGTGGGA 402
DB 631 LeuPheIleThrAspAsnMetLeuAlaLysGlnGlyPheLysAlaAsnPheThrThrGly 650
QY 403 GCTGGCTGGCGCTGGCTACTCGGAAGTTCATGGGCAGCAAGTGTCCAACCTCTGG 462
DB 651 TyrGlyLeuGly-----IleProGluProCysLysGluAspAsn 663
QY 463 ATAGAGTGGACTCTCAGGTACCTGATCAACCCCTCTAACTGCTGTGATGGCTGTCA 522
DB 664 PheGlnCys---LysAspGlyGluCysIleProLeuValAsnLeuCysAspGlyPhePro 682
QY 523 CACTCCCCGGGGGAGGAGAGATCGGTGTGTTCGCTCTACGGA----- 570
DB 683 HisCysLysAspGlySerAspGluAlaHisCysValArgLeuPheAsnGlyThrThrAsp 702
QY 571 -----CCAAACTTCATCCTTCAGGTGTACTCATCTCAGAGGAAGTCCCGCAC 618
DB 703 SerSerGlyLeuValGlnPheArgIleGln-----SerIleTrpHis 716
QY 619 CTGTGTCCAAAGACGACTGGAACGAGAACTACGGGGCGGGCGGCTCAGGAGCAATGGGC 678
DB 717 ValAlaCysAlaGluAsnTrpThrThrGlnIleSerAspAspValCysGlnLeuLeuGly 736
QY 679 TATAGAATAATTTTACTCTAGCCCAAGGAATAGTGGATGACAGCGGATCCACCACTTT 738
DB 737 LeuGlyThrGly-----AsnSerSerValProThrPheSerThrGlyGlyProTyr 754
QY 739 ATGAACATGAACACAAAGTGGCGCAATGTCGATATCTATAAATACTGTACACAGTGTAT 798
DB 755 ValAsnLeuAsnThrAlaProAsnGlySerLeuIle-----LeuThrProSerGln 771
QY 799 GCCTGTTCTTCAAAGACAGTGTCTTCTTACGCTCT-----ATAGCTCGGGGTCAAC 852
DB 799 GCCTGTTCTTCAAAGACAGTGTCTTCTTACGCTCT-----ATAGCTCGGGGTCAAC 852

```


Db 772 GlnCysLeuGluAAspSerLeuLeuLeuGlnCysAsnTyrLysSerCysGlyLysLys 791
QY 853 TTGAACCAAGCGCCAGCAGC---AGGATTCTGGCGGCGAGAGCGCGCTCCCGGGCC 909
Db 792 LeuValThrGlnGluValSerProLysIleValGlySerAspSerArgGluGlyAla 811
QY 910 TGGCCCTGGAGGTCAGCTCCAGCTCCAGAACCTCCACGCTGCGGAGGCTCCATCATC 969
Db 812 TrpProTrpValValAlaLeuTyrPheAspAspGlnGlnValCysGlyAlaSerLeuVal 831
QY 970 ACCCCGAGTGGATCGTACAGCGCCGCTGCTGGGAAACCTTTAACAATCCATGG 1029
Db 832 SerArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnMetGluProSer 851
QY 1030 CATTCGAGCATTTGGCGGGATTTTCAGACAACTTTTCATGCTTCTATGAGCGCGATAC 1089
Db 852 LysTrpLysAlaValLeuGlyLeuHisMetAlaSerAsnLeu-----ThrSerPro 868
QY 1090 CAAGTAGA-----AAAGTAGTTCTCATCCAAATTTATGACTCCAGAAC 1134
Db 869 GlnIleGluThrArgLeuLeuAAspGlnIleValIleAsnProHisTyrAsnLysArgArg 888
QY 1135 AAGAACATGATGATTCGCTGATGAGCTGAGCTGAGAGCTCTGACTTTCAACGACTAGTG 1194
Db 889 LysAsnAsnAspIleAlaMetMetHisLeuGluMetLysValAsnTyrThrAspTyrIle 908
QY 1195 AAACAGTGTCTGCTCCCAAGCCAGGCGATGATGCTGACGAGCAACAGCTCTGCTGAT 1254
Db 909 GlnProIleCysLeuProGluGluAsnGlnValPheProGlyArgIleCysSerIle 928
QY 1255 TCCGGTGGGGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1314
Db 929 AlaGlyTrpGlyAlaLeuIleTyrGlnIleSerThrAlaAspValLeuGlnGluAlaAsp 948
QY 1315 GTGCTTCTCATGACAGACAGATGCAACAGACAGATGCTATGATGCAACCTGATCACA 1374
Db 949 ValProLeuLeuSerAsnGluLysCysGlnGlnGlnMetProGluTyrAsn---IleThr 967
QY 1375 CCAGCCATGATCTGCGCGCTTCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434
Db 968 GluAsnMetValCysAlaGlyTyrGluAlaGlyValAspSerCysGlnGlyAspSer 987
QY 1435 GGAGGCGCTGCTGCTCGAAGCAACAATCTGGTGGCTGATAGGGGATACAGCTGG 1494
Db 988 GlyGlyProLeuMetCysGlnGluAsnAsnArgTrpLeuLeuAlaGlyValThrSerPhe 1007
QY 1495 GGTTCGCTGCTGCGCAAGCTTACAGACAGGATGACGAGGATGATGATGATGATGATGAT 1554
Db 1008 GlyTyrGlnCysAlaLeuProAsnArgProGlyValTyrAlaArgValProArgPheThr 1027
QY 1555 GACTGGATT 1563
Db 1028 GluTrpIle 1030

RESULT 9

TMS5_HUMAN STANDARD; PRT; 457 AA.
ID Q9H3S3;
AC Q9H3S3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
GN TMSR55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TMS5-Brain;
RX PubMed-11741986;
RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
RT "Spinesin/TMSR55, a novel transmembrane serine protease, cloned from

human spinal cord.";
J. Biol. Chem. 277:6806-6812(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).
CC -1- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
CC neurons, in their axons, and at the synapses of motoneurons in the
CC spinal cord.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AB028140; BAB20375.1; -;
DR RSP; P00763; IDPO.
DR Genew; HGNC:14908; TMSR55.
DR MIM; 606751; -;
DR MEROPS; S01.313; -;
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; peptidase S1.
DR InterPro; IPR001314; peptidase S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; P00722; CHYMOTRYPSIN.
DR SMART; SM0020; TRYP SP; 1.
DR PROSITE; PS0240; TRYP SIN DOM; 1.
DR PROSITE; PS00134; TRYP SIN HIS; 1.
DR PROSITE; PS00135; TRYP SIN SER; 1.
DR PROSITE; PS00420; SRCR_1; FALSE NEG.
DR PROSITE; PS0287; SRCR_2; FALSE NEG.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein.
FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 71 457 (POTENTIAL).
FT DOMAIN 112 207 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 218 457 SRCR.
FT ACT_SITE 258 258 SERINE PROTEASE.
FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 218 CLEAVAGE (POTENTIAL).
FT DISULFID 135 196 BY SIMILARITY.
FT DISULFID 148 206 BY SIMILARITY.
FT DISULFID 209 328 BY SIMILARITY.
FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT DISULFID 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 195 195 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;
Alignment Scores:
Pred. No.: 3,59e-39 Length: 457
Score: 651.50 Matches: 157
Percent Similarity: 46.65% Conservative: 59
Best Local Similarity: 33.91% Mismatches: 203
Query Match: 20.49% Indels: 44
DB: 1 Gaps: 10

US-09-323-597C-1 (1-1738) x TMS5_HUMAN (1-457)

QY 246 CCGTCCCGGTCGCCAGTACGCCCGGAGGTCCTGACGAGGCTTCCAAACCCGTCGT 305

Db 9 ProProMetGluAlaGlnTyrAlaGluGluGlyProGlyProGlyIlePheArgAlaGlu 28

QY 306 CTGACGCGCCAAATCCCATCCCGGACAGTGTGCACCTCAAGACTAAGAAAGCACT 365
 Db 29 ProGlyAspGlnGlnHisProIleSerGlnAlaVal-CysTrpArgSerMetArgValG 48
 QY 366 GTGCATCACCTTGACCTCGGGAGACCTCTCTCGTGGAGCTCGGCTGGCGTGCCTACT 425
 Db 48 yCysAlaValLeuGlyAlaLeuGlyLeuAlaGlyAlaGlyValGlySerTrpLeuLe 68
 QY 426 CTGAAAGTTCATGGCGACAGTGTCTCAAC-----TCTGGATAGAGTCGACTCCTC 479
 Db 68 uValLeuTyrLeuCysProAlaAlaSerGlnProIleSerGlyThrLeuGlnAspGluG 88
 QY 480 AGGTACCTGCATCAACCCCTCTAACTGGTGTGTGATGGCTGCACACTGCCCGCGGGGA 539
 Db 88 uileThrLeu-----SerCysSerGluAlaLe 97
 QY 540 GGACGAGATCGGTGTGT-----GCCCTCTACGGAC 572
 Db 97 rAlaGluGluAlaLeuLeuProAlaLeuProLysThrValSerPheArgIleAsnSerG 117
 QY 573 AAATCTCATCTCTCAGGTGTACTCATCTCAGAGGAAGTCTCGGCACCTGTGTGCCAAGA 632
 Db 117 uAspPheLeuLeuGluAlaGlnValArgAspGlnProArgTrpLeuValCysHisG 137
 QY 633 CGACTGGAACGAGAACTACGGCGCGGCGCTGCAGGACACATGGGCTAT---AAGAATAA 689
 Db 137 uGlyTrpSerProAlaLeuGlyLeuGlnIleCysTrpSerLeuGlyHisLeuArgLeuTh 157
 QY 690 TTTTACTCTAGCAAGATAGTGGATGACAGCGGATCCACAGCTTTATGAACCTGAA 749
 Db 157 rHisHisGlyValAsnLeuThrAsp-----lleIysLeuAs 170
 QY 750 CACAAGTCCCGCAATCTCATATCTATAAAAACTG-----TACCACAGTGTATGCC-- 801
 Db 170 nSerSerGlnGluPheAlaGlnLeuSerProArgLeuGlyPheLeuGluAlaTr 190
 QY 802 -----TGTTCTTCAAAGCAGTGGTTTCTTTTACCTGTGTATAGCTCGCG 845
 Db 190 pGlnProArgAsnAsnCysThrSerGlyGlnValSerLeuArgCysSerGluCysG 210
 QY 846 GGTCAACTTGAACTCAAGCCCGACAGATGTGGCGGCGAGAGCGGCTCCCGGG 905
 Db 210 yAlaArg-----ProLeuAlaSerArgIleValGlyGlyGlnSerValAlaProG 227
 QY 906 GGCTGCTCCCTGGAGTGCAGCTGCAGTCCAGAAAGTCCAGTGTGGAGGCTCCAT 965
 Db 227 yArgTrpProGlnAlaSerValAlaLeuGlyPheArgHisThrCysGlyGlySerVa 247
 QY 966 CATCACCCCGAGTGGATGTGACAGCGGCCACTGCTGGAAAAA---CCTCTTAACA 1022
 Db 247 lLeuAlaProArgTrpValThrAlaAlaHisCysMetHisSerPheArgLeuAlaAr 267
 QY 1023 TCCATGGCATTTGGACGCAATTTGGGGGATTTTGAGACAATCTTTTCATGTCTATGGAGC 1082
 Db 267 gLeuSerSerTrpArgValHisAlaGlyLeuValSerHisSerAlaValArgProHisG 287
 QY 1083 CGGATACCAAGTAGAAAAAGTATTTCTCTCCAAATATGATCTCCAGACCAAGAACAA 1142
 Db 287 nGlyAlaLeuValGluArgIleIleProHisProLeuTyrSerAlaGlnAsnHisAspTy 307
 QY 1143 TGCAATGTGCTGATGAAGTGCAGAGCGCTCTGACCTTCAACACCACTAGTGAACCACT 1202
 Db 307 rAspValAlaLeuLeuArgLeuGlnThrAlaLeuAsnPheSerAspThrValGlyAlaVa 327
 QY 1203 GTGTCTGCCCAACCCAGCGCATGTGTCGACGAGAACAGCTCTGTCTGGATTTCCGGGTG 1262
 Db 327 lCysLeuProAlaLysGluGlnHisPheProLysGlySerArgCysTrpValSerGlyTr 347
 QY 1263 GGGGGCCACCGAGGAGGAAGGAG---ACCTCAGAAGTGTCTGACCTGCCAAGTGTCT 1319
 Db 347 pGlyHisThrHisProSerHisThrTyrSerSerAspMetLeuGlnAspThrValValPr 367
 QY 1320 TCTCATTGACACAGAGATGCAACAGCAGATATGTCTATGACAAACCTGATCACACAGC 1379

Db 367 oLeuPheSerThrGlnLeuCysAsnSerSerCysValTyrSerGlyAlaLeuThrProAr 387
 QY 1380 CATGATCTGTCGGCTTCTCTCGAGGAGACGTCGATTCTTCGAGGTGACAGTGGAGG 1439
 Db 387 gMetLeuCysAlaGlyTyrLeuAspGlyArgAlaAspAlaCysGlnGlyAspSerGlyG 407
 QY 1440 GCCTCTGGTCTCCTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAAAGTGGGTTTC 1499
 Db 407 yProLeuValCysProAspGlyAspThrTrpArgLeuValGlyValValSerTrpGlyAr 427
 QY 1500 TGCGTGTGCCAAAGCTTACAGACACGAGGTGTACGGGAATGTGATGTATTCACGGAAGT 1559
 Db 427 gAlaCysAlaGluProAsnHisProGlyValTyrAlaLysValAlaGluPheLeuAspTr 447
 QY 1560 GATTAT 1566
 Db 447 pIleHis 449
 RESULT 10
 ENTK_HUMAN
 ID ENTK_HUMAN STANDARD; PRT; 1019 AA.
 AC P98073;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Enterokinase precursor (EC 3.4.21.9) (Enterokinase).
 GN PRSS7 OR ENTK
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ductenum;
 RX MEDLINE=95234679; PubMed=7718557;
 RA Kitamoto Y., Veille R.A., Donis-Keller H., Sadler J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 RT the proteolytic activator of trypsinogen.";
 RL Biochemistry 34:4562-4568(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND DISEASE.
 RX MEDLINE=21606074; PubMed=11719902;
 RA Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kapplier M.,
 RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;
 RT "Mutations in the proenteropeptidase gene are the molecular cause of
 RT congenital enteropeptidase deficiency.";
 RL Am. J. Hum. Genet. 70:20-25(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Dalabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine G., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramer J., Beck A., Klages S., Kennig S., Rieselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Leinhardt H., Reinhardt R., Vasp M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [4]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE=Ductenum;
 RX MEDLINE=94329561; PubMed=8052624;
 RA Kitamoto Y., Yuan Q., Wu Q., McCourt D.W., Sadler J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 RT protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).

FT	CONFLICT	754	771	SQGLQDSLRQLQCNHKS -> RENAKNEIDALSPILLIA (IN REF. 3).	
Alignment Scores:					
Pred. No.:	8.93e-39	Length:	1019		
Score:	647.00	Matches:	138		
Percent Similarity:	53.25%	Conservative:	67		
Best Local Similarity:	35.84%	Mismatches:	154		
Query Match:	20.35%	Indels:	26		
DB:	1	Gaps:	10		
US-09-323-597C-1 (1-1738) x ENT_K_HUMAN (1-1019)					
QY	448	TGCTCCAACTCTGGGATAGTGGAGCTCTCAGTACCTGCATCAACCCCTCTAATCGG	507		
DB	643	CyslysalasphHisPheGlnCys---LysasnGlyGluCysValProLeuValAsnLeu	661		
QY	508	TGTGATGGCGTGCACATGCTCCCGGGGGAGGACGAGATCGGTGTGTCGCTCTAC	567		
DB	662	CysAspGlyHisLeuHisCysGluAspGlySerAspGluAlaAspCysValArgPhe	681		
QY	568	GGA-----CCAACTTCATCTCTCAGGTGACTCATCTCAG	603		
DB	682	AsnGlyThrThrAsnAsnAsnGlyLeuValArgPheArgileGln-----	696		
QY	604	AGGAAGTCTCGCCCTGCTGTCGAAGACGACTGGAACGAGAACTACGGCGCGGCC	663		
DB	697	---SerileThrHisThrAlaCysAlaGluAsnTrpThrThrGlnIleSerAsnAspVal	715		
QY	664	TGCAGGACATGGGCTATAAATAATTTTACTCTAGCCCAAGCAATAGTGGATCAGC	723		
DB	716	CysGlnLeuLeuGlyLeuGlySerGly---AsnSerSerLysProIlePheSerThrAsp	734		
QY	724	GGATCCACACGCTTTATGAACACTGAACACACAGTCCGCGCAATGCGATATCATAAAAA	783		
DB	735	GlyGly---ProPheValLysLeuAsnThrAlaPro-----AspGlyHisLeuIle	750		
QY	784	CTGTACACAGTGTGCTGCTGTTCTTCAACACAGTGTGTTCTTACCGCTG-----ATA	837		
DB	751	LeuThrProSerGlnGlnCysLeuGlnAspSerLeuIleArgLeuGlnCysAsnHisLys	770		
QY	838	GCCTCGCGGGTCAACTGCACTCAAGCGC---CAGACGAGATGTGGGGGGGAGAGC	894		
DB	771	SerCysGlyLysLysLeuAlaAlaGlnAspIleThrProLysIleValGlyGlySerAsn	790		
QY	995	CGCTCCCGGGGGCTGCGCTGCGAGTGTGCTGCGAGTCCAGACGTCACGTCGTC	954		
DB	791	AlaLysGluGlyAlaTrpProTrpValValGlyLeuTyTyTyGlyGlyArgLeuLeuCys	810		
QY	955	GGAGGCTCCATCATCACCACCCCGAGTGTGATCGTGCAGCGCCACTCGGTGAAAAACCT	1014		
DB	811	GlyAlaSerLeuValSerSerAspTrpLeuValSerAlaAlaHisCysValTyArg	830		
QY	1015	CTTAACAATCCATGCGATTCGAGCGCATTTCCGGGGATTTTGACACAACTCTTCATGTC	1074		
DB	831	AsnLeuGluProSerLysTrpThrAlaIleLeuGlyLeuHisMetLysSerAsnLeuThr	850		
QY	1075	-----TATGGAGCGGATACCAAGTATGATAAAGTATTTCTCATCCAAATATGATCC	1128		
DB	851	SerProGlnThrValProArgLeuIleAspGluIleValIleAsnProHisTyArg	870		
QY	1129	AAGACCAAGAACATGATTCGCTGATGAAGTGCAGAGCGCTCTGACTTTCAACGAC	1188		
DB	871	ArgArgLysAspAsnAspIleAlaMetMetHisLeuGluPheLysValAsnTyThrAsp	890		
QY	1189	CTAGTAAACCAAGTGTGTGCTGCCCAACCCAGGATGATGCTGCGAGCAGACAGCTGTC	1248		
DB	891	TyrIleGlnProIleCysLeuProGluGluAsnGlnValPheProArgLysArgAsnCys	910		
QY	1249	TGGATTTCCGGTGGGGCCACCGAGGAGAAAGAGAGCTCAGAAGTGTCAACGCT	1308		
DB	911	SerIleAlaGlyTrpGlyThrValValTyArgGlnGlyThrThrAlaAsnIleLeuGlnGlu	930		

QY	1309	GCCAAAGTGCTTCTCATTGAGACACAGATGCAACAGCAGATATCTCTATGACAACCTG	1368		
DB	931	AlaAspValProLeuLeuSerAsnGluArgCysGlnGlnMetProGluTyArgAsn---	949		
QY	1369	ATCACACCAAGCCATGATCTGTGCGGCTCTCTCAGGGAACCTGATTTCTTCCAGGCT	1428		
DB	950	IleThrGluAsnMetIleCysAlaGlyTyArgGluGluGlyGlyIleAspSerCysGlnGly	969		
QY	1429	GACAGTGGAGGGCTCTGCTCACTTCGAGAACAAATATCTGGTGGCTGATAGGGATACA	1488		
DB	970	AspSerGlyGlyProLeuMetCysGlnGluAsnAsnArgTrpPheLeuAlaGlyValThr	989		
QY	1489	AGCTGGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1548		
DB	990	SerPheGlyTyLysCysAlaLeuProAsnArgProGlyValTyAlaArgValSerArg	1009		
QY	1549	TTCACGAGCTGGATT	1563		
DB	1010	PheThrGluTrpIle	1014		
RESULT 11					
ENT_K_MOUSE	ENT_K_MOUSE	STANDARD;	PRT;	1069	AA.
ID	P97435;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Enteropeptidase (EC 3.4.21.9) (Enterokinase).				
GN	PRSS7 OR ENT_K.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NCBI	TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STEAIN=C57BL/6; TISSUE=DuoDenum;				
RX	MEDLINE=9817142; PubMed=9486188;				
RA	yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;				
RT	"Structure of murine enterokinase (enteropeptidase) and expression in				
RT	small intestine during development."				
RL	Am. J. Physiol. 274:G342-G349(1998).				
CC	-I- FUNCTION: Responsible for initiating activation of pancreatic				
CC	proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase				
CC	A). It catalyzes the conversion of trypsinogen to trypsin which in				
CC	turn activates other proenzymes including chymotrypsinogen,				
CC	procarboxypeptidases, and proelastases (By similarity).				
CC	-I- CATALYTIC ACTIVITY: Selective cleavage of 6-lys-- -lle-7 bond in				
CC	trypsinogen.				
CC	-I- SUBUNIT: Heterodimer of a catalytic (light) chain and a				
CC	multidomain (heavy) chain linked by a disulfide bond (By				
CC	similarity).				
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein (Probable).				
CC	-I- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS				
CC	CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).				
CC	-I- SIMILARITY: Belongs to peptidase family S1.				
CC	-I- SIMILARITY: Contains 2 CUB domains.				
CC	-I- SIMILARITY: Contains 2 LDL-receptor class A domains.				
CC	-I- SIMILARITY: Contains 1 KAM domain.				
CC	-I- SIMILARITY: Contains 1 SEA domain.				
CC	-I- SIMILARITY: Contains 1 SRCR domain.				
CC	-I- SIMILARITY: Contains 1 SRCR domain.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	EMBL; U73378; AAB37317.1; --				
DR	HSPSP; Q07954; ICR8.				
DR	MEKOPS; S01156; --				
DR	MGD; MG1:1197523; Prss7.				

FT	CHAIN	800	1034	CATALYTIC L CHAIN (LIGHT CHAIN).
FT	DOMAIN	1	18	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	19	47	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	DOMAIN	48	1034	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	52	169	SEA.
FT	DOMAIN	197	238	LDL-RECEPTOR CLASS A 1.
FT	DOMAIN	240	349	CUB 1.
FT	DOMAIN	357	519	NAM.
FT	DOMAIN	539	649	CUB 2.
FT	DOMAIN	656	694	LDL-RECEPTOR CLASS A 2.
FT	DOMAIN	693	786	SRCR.
FT	DOMAIN	800	1034	SERINE PROTEASE.
FT	ACT_SITE	840	840	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	891	891	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	986	986	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	N-myristoyl glycine (Potential).
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	236	BY SIMILARITY.
FT	DISULFID	658	670	BY SIMILARITY.
FT	DISULFID	665	683	BY SIMILARITY.
FT	DISULFID	677	692	BY SIMILARITY.
FT	DISULFID	787	911	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	825	841	BY SIMILARITY.
FT	DISULFID	925	932	BY SIMILARITY.
FT	DISULFID	956	971	BY SIMILARITY.
FT	DISULFID	982	1010	BY SIMILARITY.
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	170	170	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	194	194	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	283	283	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	343	343	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	403	403	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	455	455	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	518	518	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	549	549	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	697	697	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	701	701	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	721	721	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	740	740	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	761	761	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	804	804	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	863	863	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	1034	AA, 114776 MW, 0388C64CF64CC368 CRC64;	

Alignment Scores:
Pred. No.: 7,67e-38 Length: 1034
Score: 634.00 Matches: 147
Percent Similarity: 49.58% Conservative: 72
Best Local Similarity: 33.28% Mismatches: 179
Query Match: 19.94% Indels: 44
DB: 1 Gaps: 13

US-09-323-597C-1 (1-1738) x ENT_K_PIG (1-1034)

QY	304	GTCTGCACGACGCCCAATCCCATCCCGGACATGTGCACCTCAAGACTAAGAAGCA	363
Db	610	ValTyThrGlyProGlyProValGluAspValPheSerThrThrAsnArgMetThrVal	629
QY	364	CTGTGTCATCACCC-----TTGACCTCGGG-----ACCTCTCGTGGGA	402
Db	630	LeupPheIleThrAsnAspAlaLeuThrIysGlyGlyPheIysAlaAsnPheThrThrGly	649
QY	403	GCTGCCTGCGCGCTGCGCTACTCTGGAAGTTCAAGGCGACGACGAGTGCTCCAACTCGGG	462
Db	650	TyrHisLeuGly-----IleProGluProCysIysGluAspAsn	662

Alignment Scores:

Pred. No.: 7, 67e-38 Length: 1034
 Score: 634.00 Matches: 147
 Percent Similarity: 49.53% Conservative: 72
 Best Local Similarity: 33.28% Mismatches: 179
 Query Match: 19.94% Indels: 44
 DBs: 1 Gaps: 13

US-09-323-597c-1 (1-1738) x ENT_K_FIG (1-1034)

QY	304	GTCTGACGACGACCCCAATCCCATCCGGGACAGTGTGCACCTCAAGACTAAGAACCA	363
DB	610	ValTyrThrGlyProGlyProValGluAspValPheSerThrThrAsnArgMetThrVal	629
QY	364	CTGTGCATCACCC-----TTGACCCCTGGGG-----ACCTTCCTCGTGGGA	402
DB	630	LeuPheIleThrAsnAspAlaLeuThrLysGlyGlyPheLysAlaAsnPheThrThrGly	649
QY	403	GCTGCGTGGCGCTGCGCTACTCTGTGAAGTTCATGCGGACGACCAAGTGTCCAACTCTGGG	462
DB	650	TyrHisLeuGly-----11eProGluProCysLysGluAspAsn	662

QY	463	ATAGAGTGGCACTCCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGTGATGGGTGTCA	522
DB	663	PheGlnCysGlu---AsnGlyGluCysValLeuValAsnLeuCysAspGlyPheSer	681
QY	523	CACCTGCCCCGGGGAGGAGCAGAAATCGGTGTGTGTTCCCTCTACGGA-----CCAAC	576
DB	682	HisCysLysAspGlySerAspGluAlaHisCysValArgPheLeuAsnGlyThrAlaAsn	701
QY	577	TTCATCTCCTCAGGTGTACTCATCTCAGAGGAAGAGCTCGGCACCCCTGTGTGCAAGACAC	636
DB	702	AsnSerGlyLeuValGlnPheArgGlnSerIleThrAlaCysAlaGluAsn	721
QY	637	TGGAACGAGAACTACGGCGCGGCTCGACGGACATGGGCTATAGATAAT-----	690
DB	722	TrpThrThrGlnThrSerAspValCysGlnLeuLeuGlyLeuGlyThrGlyAsnSer	741
QY	691	-----TTTTACTCTAGCCCAAGGAATAGTGTGATGACAGCGGCTCACCAGCTTTATG	741
DB	742	SerMetProPhePheSer-----SerGlyGlyGlyProPheVal	754
QY	742	AACTGACACACAGTCCCGCAATGTGTATATATAAACTGTACCAAGTATGCC	801
DB	755	LysLeuAsnThrAlaProAsnGlySerLeuIle-----LeuThrAlaSerGluGln	771
QY	802	TGTTCTTCAAAAGCAGTGTTCCTTTACGCTGT-----ATAGCCTCGGGGTCAACTG	855
DB	772	CysPheGluAspSerLeuIleLeuLeuGlnCysAsnHisLysSerCysGlyLysGln	791
QY	856	AACTCAAGCCCGCAGAC---AGGATTGTGGGGGGGAGCGCGCTCCCGGGGCGCTGG	912
DB	792	ValAlaGlnGluValSerProLysIleValGlyAsnAspSerArgGluGlyAlaTrp	811
QY	913	CCCTGGCAGCTCAGCTGCAGCTCCAGAACGTCACGTCGTGTGGAGGCTCCATCATCACC	972
DB	812	ProTrpValValAlaLeuTyrTyrAsnGlyGlnLeuLeuCysGlyAlaSerLeuValSer	831
QY	973	CCCGAGTGGATCGTGACAGCGCCCACTGGGTGGAAAAACCTCTTAACAATCCATGGCAT	1032
DB	832	ArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnLeuGluProSerLys	851
QY	1033	TGGACGGCATTTGGCGGGATTTGAGACATCTTCATGTTCTATGAGCGCGGTATACAA	1092
DB	852	TrpLysAlaIleLeuGlyLeuHisMetThrSerAsnLeu-----ThrSerProGln	869
QY	1093	-----GTAGAAAAGTGTATCTCATCCAAATATGACTCAAGACCAAG	1137
DB	869	IleValThrArgLeuIleAspGluIleValIleAsnProHisTyrAsnArgArgLys	888
QY	1138	AACAATGACATTTGGCTGTGATGAAGCTGCAGAGCCTCTGACTTTCAACGACCTAGTAAA	1197
DB	889	AspSerAspIleAlaMetMetHisLeuGluPheLysValAsnTyrThrAspTyrIleGln	908
QY	1198	CCAGTGTGTCTGCCACCCAGCATGCTGCAGCCAGAAACAGCTCTCTGTGATTCC	1257
DB	909	ProIleCysLeuProGluGluAsnGlnValPheProProGlyArgIleCysSerIleAla	928
QY	1258	GGGTGGGGGGCCCGAGGAGAAAGGAGACCTCAGAACTGTGTGAACGCTGCCAAGGTG	1317
DB	929	GlyTrpGlyLysValIleTyrGlnGlySerProAlaAspIleLeuGlnGluAlaAspVal	948
QY	1318	CTTCTCATTGACACACAGATGTCACAGCAGATATGCTATGACACACCTGATCACCA	1377
DB	949	ProLeuLeuSerAsnGluLysCysGlnGlnMetProGluTyrAsn---IleThrGlu	967
QY	1378	GCCATGATCTGTCCGGCTTCCCGAGGGAAGCTCGAATTTGCCAGGTGTGACGTGGA	1437
DB	968	AsnMetMetCysAlaGlyTyrGluGluGlyGlyIleAspSerCysGlnGlyAspSerGly	987
QY	1438	GCGCTTGTCTGCTTCGAGACATATCTGTGGCTGTAGTGGGATACAGCTCGGGT	1497
DB	988	GlyProLeuMetCysLeuGluAsnAspArgTrpLeuLeuAlaGlyValThrSerPheGly	1007

Qy 1498 TCTGGCTGTGCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTTCACGGAC 1557
 Db 1008 TyrcGlnCysAlaLeuProAsnArgProGlyValTyralaArgValProLysPheThrGlu 1027
 Qy 1558 TGGATT 1563
 Db 1028 Tpile 1029
 RESULT 13
 CORI_HUMAN
 ID CORI_HUMAN STANDARD; PRT; 1042 AA.
 AC Q9Y5Q3; Q9UHY2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
 DE converting enzyme) (Corin) (Heart specific serine proteinase ATC2).
 GN CRN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=99262646; PubMed=10329693;
 RA Yan W., Sheng N., Seto M., Morser J., Wu Q.;
 RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
 RT from human heart.";
 RL J. Biol. Chem. 274:14926-14935(1999).
 RN [2]
 RP SEQUENCE OF 734-1040 FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=20534769; PubMed=11082206;
 RA Hooper J.D., Scarman A.L., Clarke B.E., Normyle J.F., Antalis T.M.;
 RT "Localization of the mosaic transmembrane serine protease corin to
 RT heart myocytes.";
 RL Eur. J. Biochem. 267:6931-6937(2000).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=20359740; PubMed=10880574;
 RA Yan W., Wu F., Morser J., Wu Q.;
 RT "Corin, a transmembrane cardiac serine protease, acts as a pro-atrial
 RT natriuretic peptide-converting enzyme.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:8525-8529(2000).
 CC -!- FUNCTION: Converts Pro-ANP to ANP. Cleaves Pro-ANP specifically
 CC between Arg-123 and Ser-124.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -!- TISSUE SPECIFICITY: Highly expressed in heart. Expressed in heart
 CC myocytes.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 2 frizzled (FZ) domains.
 CC -!- SIMILARITY: Contains 7 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF113845; AAD31850.1; --
 DR EMBL; AF113248; AAP21966.1; --
 DR HSPSP; P00763; 1DPO.
 DR MEROPS; S01.019; --
 DR MIN; G05236; --
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0004232; F:serine-type endopeptidase activity; TAS.
 DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
 DR GO; GO:0006629; P:lipid metabolism; TAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.

DR GO; GO:0008217; P:regulation of blood pressure; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR000024; Fz_domain.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF01392; Fz_2.
 DR Pfam; PF00057; ldl_recept_a; 6.
 DR Pfam; PF00089; trypsin_1.
 DR PRINTS; PR00262; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00063; FRI; 2.
 DR SMART; SM00192; LDLa; 3.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Ttyp_spc; 1.
 DR PROSITE; PS00038; FZ_2.
 DR PROSITE; PS01209; LDLRA_1; 6.
 DR PROSITE; PS00068; LDLRA_2; 7.
 DR PROSITE; PS0240; TRYPsin DOM; 1.
 DR PROSITE; PS0134; TRYPsin HIS; FALSE_NEG.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; FALSE_NEG.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Repeat.
 FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 46 66 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT FT (POTENTIAL).
 FT FT EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 67 1042
 FT DOMAIN 134 259
 FT DOMAIN 268 304
 FT DOMAIN 305 340
 FT DOMAIN 341 377
 FT DOMAIN 378 415
 FT DOMAIN 450 573
 FT DOMAIN 579 614
 FT DOMAIN 615 653
 FT DOMAIN 654 690
 FT DOMAIN 690 786
 FT DOMAIN 802 1042
 FT ACT_SITE 843 843 SERINE PROTEASE.
 FT ACT_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 985 985 CHARGE RELAY SYSTEM.
 FT DISULFID 790 912 BY SIMILARITY.
 FT DISULFID 828 844 BY SIMILARITY.
 FT DISULFID 955 970 BY SIMILARITY.
 FT DISULFID 981 1010 BY SIMILARITY.
 FT CARBOHYD 80 80 N-LINKED (GLCNAC...)
 FT CARBOHYD 104 104 N-LINKED (GLCNAC...)
 FT CARBOHYD 135 135 N-LINKED (GLCNAC...)
 FT CARBOHYD 141 141 N-LINKED (GLCNAC...)
 FT CARBOHYD 231 231 N-LINKED (GLCNAC...)
 FT CARBOHYD 245 245 N-LINKED (GLCNAC...)
 FT CARBOHYD 251 251 N-LINKED (GLCNAC...)
 FT CARBOHYD 305 305 N-LINKED (GLCNAC...)
 FT CARBOHYD 320 320 N-LINKED (GLCNAC...)
 FT CARBOHYD 376 376 N-LINKED (GLCNAC...)
 FT CARBOHYD 413 413 N-LINKED (GLCNAC...)
 FT CARBOHYD 446 446 N-LINKED (GLCNAC...)
 FT CARBOHYD 451 451 N-LINKED (GLCNAC...)
 FT CARBOHYD 469 469 N-LINKED (GLCNAC...)
 FT CARBOHYD 567 567 N-LINKED (GLCNAC...)
 FT CARBOHYD 651 651 N-LINKED (GLCNAC...)
 FT CARBOHYD 697 697 N-LINKED (GLCNAC...)
 FT CARBOHYD 761 761 N-LINKED (GLCNAC...)
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC...)
 FT MUTAGEN 985 985 S->A: LOSS OF ACTIVITY.
 FT CONFLICT 854 854 W -> R (IN REF. 2).
 FT CONFLICT 876 876 K -> R (IN REF. 2).
 SQ SEQUENCE 1042 AA; 116564 MW; 7705398EBB607AD2 CRC64;

Alignment Scores:

Pred. No.:	1,79e-35	Length:	1042
Score:	601.00	Matches:	138
Percent Similarity:	51.15%	Conservative:	62
Best Local Similarity:	35.29%	Mismatches:	169
Query Match:	18.91%	Indels:	22
DB:	1	Gaps:	14

US-09-323-597C-1 (1-1738) x CORI_HUMAN (1-1042)

QY	442	AGCAAGTCTCCAACTCTGGGATGAGCGGCACTCTCCAGGTACTCTGCATCAACCCCTCT	501
DB	653	SerPheCysGlnAspAspGluLeuGluCys---AlaAsnHisAlaCysValSerArgAsp	671
QY	502	AACGTGCTGTGAGTGGCTGTACACTCTGCCCGCGCGGAGAGACAGAGATCGGTGTTCGC	561
DB	672	LeuTrpCysAspGlyGluAlaAspCysSerAspSerSerSerGluTrpAspCysValThr	691
QY	562	CTC-----TACGGACCAAACTTCATCTCAGGTGACTCATCTCAGAGGAAGTCC	612
DB	692	LeuSerIleAenValAenSerSerPheLeuMetValHisArgAlaAlaThrGlu---	710
QY	613	TGGCACTCTGTGTGCAACAGACTGGAACAGAACTACCGCGCGCGCTCGCAGGAC	672
DB	711	--HisHisValCysAlaAspGlyTrpGlnGluIleLeuSerGlnLeuAlaCysGln	729
QY	673	ATGGCTATAGATAATTTTACTCTAGCCAAAGATAGTGGATGACAGCGGATCCACC	732
DB	730	MetGlyLeuGlyGluProSerValThrLysLeu---IleGlnGluGlnGluLysGluPro	748
QY	733	AGCTTTATGAACATGAACACA-----AGTCGGCGCAATGTCGATATCTATAAAAAA	783
DB	749	ArgTrpLeuLeuHisSerAsnTrpGluSerLeuAsnGlyThrThrLeuHisGluLeu	768
QY	784	CTGTACCAACAGTATGCTCTTCTTCAAAAGACGTGGTCTTTACGCTGTATAGCC---	840
DB	769	LeuValAsnGlyGlnSerCysGluSerArgSerLysIleSerLeuLeuCysThrLysGln	788
QY	841	---TCGGGGTCAACTCACTCAAGCCGACGAGGATGTTGGCGGCGAGGCGCG	897
DB	789	AspCysGlyArgArgProAlaAlaArgMetAsnLysArgIleLeuGlyGlyArgThrSer	808
QY	898	CTCCGGGGGCTGGCCCTCGCAGGTGACCTGACGTCAGCAG---AAGTCACGTGGTC	954
DB	809	ArgProGlyArgTrpProTrpGlnCysSerLeuGlnSerGluProSerGlyHisIleCys	828
QY	955	GGAGCTCCATCATACCCCGAGTGGATCTGTGACGCGCCGACCTGCTGGGAAACCT	1014
DB	829	GlyCysValLeuIleAlaLysLysTrpValLeuThrValAlaHisCysPheGlu---Gly	847
QY	1015	CTTAACTCAATGATGCGCTGACGCGCATTTGCGGGCATT-----TTGAGCAATCTTTC	1068
DB	848	ArgGluAsnAlaAlaValTrpLysValValLeuGlyIleAsnAsnLeuAspHisProSer	867
QY	1069	ATGTTCTATGGACCGGATACCAAGTAGAAAAGTATTCTTCFCCAATATGACTCC	1128
DB	868	ValPheMetGlnThrArgPhe---ValLysThrIleLeuHisProArgTrpSerArg	886
QY	1129	AAGACCAAGAACATGACATGCGCTGATGAGCTGCAGAGAGCTCTCAGCTTCAACGAC	1188
DB	887	AlaValAlaAspTrpAspIleSerIleValGluLeuSerGluAspIleSerGluThrGly	906
QY	1189	CTAGTAAACCACTGTGTCTCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTGTC	1248
DB	907	TyrValArgProValCysLeuProAsnProGluGlnTrpLeuGluProAspThrTyrCys	926
QY	1249	TGGATTCCGGTGGGGGCCACCGAGGAAAGGAGACCTCAGAGAGTGTCAACGCT	1308
DB	927	TyrIleThrGlyTrpGly-----HisMetGlyAsnLysMetProPheLysLeuGlnGlu	944
QY	1309	GCCAAAGGTCTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTCTATGACAACCTG	1368
DB	945	GlyGluValArgIleIleSerLeuGluHisCysGlnSer---TyrPheAspMetLysThr	963

RESULT 14

CORI_MOUSE STANDARD; PRT; 1113 AA.

ID CORI_MOUSE AC Q92319;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-converting enzyme) (Corin) (low density lipoprotein receptor related protein 4).

DE CRN OR LRP4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

XP [1]

SEQUENCE FROM N.A.

RA MEDLINE=98429596; PubMed=9756524;

RA Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;

RT "A novel low-density lipoprotein receptor-related protein with type II membrane protein-like structure is abundant in heart.";

RL J. Biochem. 124:784-789 (1998).

CC -1- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY BETWEEN ARG-122 AND SER-123 (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Type II membrane protein.

CC -1- TISSUE SPECIFICITY: Highly expressed in heart.

CC -1- SIMILARITY: Belongs to peptidase family S1.

CC -1- SIMILARITY: Contains 2 frizzled (FZ) domains.

CC -1- SIMILARITY: Contains 7 LDL-receptor class A domains.

CC -1- SIMILARITY: Contains 1 SRCR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB013874; BAA34371.1; -

PIR; J0315; J0315.

HSP; P00763; IDPO.

MEROPS; S01.019; -

MED; MGI:1349451; LRP4.

InterPro; IPR009003; Cys Ser trypsin.

InterPro; IPR000024; Fz domain.

InterPro; IPR002172; LDL receptor A.

InterPro; IPR001254; Peptidase S1.

InterPro; IPR001314; Peptidase S1A.

InterPro; IPR001190; Srcr_receptor.

Pfam; PF01392; Fz; 2.

Pfam; PF00057; ldl_recept_a; 6.

Pfam; PF00089; trypsin; 1.

PRINTS; PR00722; CHYMOTRYPSIN.	
PRINTS; PR00261; LDLRECEPTOR.	
SMART; SM00063; FRI; 2.	
SMART; SM00192; LDLa; 3.	

```

DR SMART; SM00202; SR; 1.
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS00038; FZ; 2.
DR PROSITE; PS01209; LDLRA_1; 6.
DR PROSITE; PS00068; LDLRA_2; 7.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Repeat.
FT DOMAIN 1 112
FT TRANSMEM 113 133
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 134 113
FT DOMAIN 201 327
FT DOMAIN 336 372
FT DOMAIN 373 408
FT DOMAIN 409 445
FT DOMAIN 446 483
FT DOMAIN 518 641
FT DOMAIN 647 682
FT DOMAIN 683 721
FT DOMAIN 722 757
FT DOMAIN 758 853
FT DOMAIN 869 1113
FT ACT_SITE 910 910
FT ACT_SITE 959 959
FT ACT_SITE 1052 1052
FT DISULFID 857 979
FT DISULFID 895 911
FT DISULFID 1022 1037
FT DISULFID 1048 1077
FT CARBOHYD 147 147
FT CARBOHYD 202 202
FT CARBOHYD 208 208
FT CARBOHYD 298 298
FT CARBOHYD 317 317
FT CARBOHYD 373 373
FT CARBOHYD 411 411
FT CARBOHYD 444 444
FT CARBOHYD 481 481
FT CARBOHYD 519 519
FT CARBOHYD 537 537
FT CARBOHYD 635 635
FT CARBOHYD 719 719
FT CARBOHYD 765 765
FT CARBOHYD 828 828
FT CARBOHYD 970 970
FT CARBOHYD 1089 1089
SQ SEQUENCE 1113 AA; 122984 MW; B84532C5P20DD8EC CRC64;

Alignment Scores:
Pred. No.: 6 88e-34 Length: 1113
Score: 579.00 Matches: 138
Percent Similarity: 48.58% Conservative: 68
Best Local Similarity: 32.55% Mismatches: 162
Query Match: 18.21% Indels: 56
DB: 1 Gaps: 15

US-09-323-597C-1 (1-1738) x CORI_MOUSE (1-1113)
QY 442 AGCAAGTCTCCAACTCTGGGATAGAGTGAGTCTCTAGGTACTGCTGATCAACCCCTCT 501
DB 721 SerPheCysGlnAspAsnGluLeuGluCys---AlaAsnHisGluCysValProArgAsp 739
QY 502 AACTGGTGTGATCGCGTGTACACTGCCCGGGGGAGAGCAGAGAATCGGTGTGTGCG 561
DB 740 LeuTrpCysAspGlyTrpValaspCysSerAspSerSerAspGluTrpGlyCysValThr 759
QY 562 CTCTAC-----GGACCAAACTTCATCTTTCAGGTGTACTCATCTCAGAGAGTCC 612
DB 1113

```

```

Db 760 LeuSerLysAsnGlyAsnSerSerSerLeuLeuThrValHisLysSerAlaLysGlu--- 778
QY 613 TGGCAACCTGTGTGCGCAAGACGACTGGAAACGAGAACTACGGCGCGCGCGCTGCGAGGAC 672
DB 779 ---HisHisValCysAlaAspGlyTrpArgGluThrLeuSerGlnLeuAlaCysLysGln 797
QY 673 ATGGGCTATAGAATAATTTTACTCTAGCAAGAAAGTAAGTAGTACAGACGGATCCACC 732
DB 798 MetGlyLeuGlyGluPro-----SerValThrLysLeuLeuProGlyGlnGluGln 815
QY 733 ACCTTTATGAACACTGAACCAAGTCCCGCAATGTCGAT-----ATCTATATAAAA 783
DB 816 GlnTrpLeuArgLeuTyProAsnTrpGluAsnLeuAsnGlySerThrLeuGlnGluLeu 835
QY 784 CTGTACCAACAGATGCGCTGCTTCAAAAGCAGGTGTTCTTTTACGCTGTATAGCC---- 840
DB 836 LeuValTyArgHisSerCysProSerArgSerGluLeuSerLeuLeuCysSerLysGln 855
QY 841 ---TCGGGGTCAACTTCAACCTCAAGCCGACAGAGATGTGGCGCGCGAGACGCG 897
DB 856 AspCysGlyArgArgProAlaAlaArgMetAsnLysArgGlyLeuGlyGlyArgThrSer 875
QY 898 CTCGGGGGGCTGCGCTGCGAGCTGACGCTGACGTCAGCTCCAG---AACGTCCACGTGTC 954
DB 876 ArgProGlyArgTrpTrpGlnCysSerLeuGlnSerGluProSerGlyHisLysCys 895
QY 955 GGAGCTCCATCATCACCCCGAGTGGATCGTACAGCGCGCCACTCGGTGCAA----- 1008
DB 896 GlyCysValLeuLeuAlaLysLysTrpValLeuThrValAlaHisCysPheGluGlyArg 915
QY 1009 -----AAACCTCTTAAACAATCCATGGCAT 1032
DB 916 GluAspAlaAspValTrpLysValValPheGlyLeuAsnAsnLeuAspHisPro----- 933
QY 1033 TGGACGGCATTTGCGGGATTTTGAGACAATCTTTCATGTTCTATGGAGCCGGATACCAA 1092
DB 934 -----SerGlyPheMetGlnThrArgPhe----- 941
QY 1093 CTAGAAAAGTGATTTCTCTCAAAATTAATGACTCTCAAGACCAAGAACAAATGACATTCGG 1152
DB 942 ValLysThrLeuLeuHisProArgTrpSerArgAlaValAlaValAspTrpAspLysSer 961
QY 1153 CTGATGAGCTGCAGAACCTCTGACTTTCACAGCCTAGTGAACACGAGTGTCTGCC 1212
DB 962 ValValGluLeuSerAspAspLeuGluThrSerTyValArgProValCysLeuPro 981
QY 1213 AACCCAGCAGTATGCTGCGACCAAGACAGCTCTGCTGATTCGCGTGGGGGCCACC 1272
DB 982 SerProGluGluTyLeuGluProAspThrTyCysTyThrLeuThrGlyTrpGly----- 999
QY 1273 GAGGAGAAAGGAGACACTCAAGTGTGTAAGCTGCCAAGGTCTCTCTCATTCAGACA 1332
DB 1000 HisMetGlyAsnLysMetProPheLysLeuGlnGluGlyGluValArgLeuLeuProLeu 1019
QY 1333 CAGAGATCAACAGCAGATATGCTATGACAACCTGATCACACGACCATGATCTGTGCC 1392
DB 1020 GluGlnCysGlnSer---TyPheAspMetLysThrThrLeuThrAspArgMetLeuLysAla 1038
QY 1393 CGCTTCCTGCGAGGGAACGTGATTTCTTCCAGGGTGAAGTGGAGGCCCTCTGTCTACT 1452
DB 1039 GlyTyGluSerGlyThrValAspSerCysMetGlyAspSerGlyProLeuValCys 1058
QY 1453 TCGAAG---AACATATCTGCTGCTGATAGGGATACAAGCTGGGTTCTGGCTGT--- 1506
DB 1059 GluArgProGlyGlyGlnTrpThrLeuPheGlyLeuThrSerTrpGlySerValCysPhe 1078
QY 1507 GCCAAGCTTACAGACGAGGAGTGTACGGGAATGTGATGATGATTCACCGACTGGATTAT 1566
DB 1079 SerLysValLeuGlyProGlyValTySerAsnValSerTyPheValGlyTrpLeuGlu 1098
QY 1567 CGAACAATGAGGAGAGCGGCTAATCCACATGGTCTCTGCTTTCAGCTGCTTTTACAG 1626
DB 1099 ArgGlnIleTyThrLeuGln-----ThrPheLeuGlnLys 1109

```


FT DOMAIN 615 854 SERINE PROTEASE.
 FT ACT SITE 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 805 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 327 FEA -> GTR (IN REF. 5; AAH05826).
 FT CONFLICT 381 R -> S (IN REF. 4).
 FT CONFLICT 674 A -> V (IN REF. 3).
 SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Alignment Scores:
 Pred. No.: 1,16e-33 Length: 855
 Score: 575.50 Matches: 168
 Percent Similarity: 43.96% Conservative: 72
 Best Local Similarity: 30.77% Mismatches: 190
 Query Match: 18.10% Indels: 116
 DB: 1 Gaps: 23

US-09-323-597C-1 (1-1738) x ST14_HUMAN (1-855)

Qy	186	CTATCCGACAGCCCACTGGTCCCACTGTCTACAGAGTGATCCGGCTCAGTACTA	245
Db	334	LeuProArgMetSerSerCysGlyArg---LeuArgLysAlaGlnGlyThrPhe---	351
Qy	246	CCCGTCCCGCTGCCAGTACGCCCGAGGGTCTCAGCAGGCTCCACCCCGTGT	305
Db	352	-----AsnSerProTyrTrpPro-----GlyHisTyrProProAsn	363
Qy	306	CTGCACGACGCC---CRAATCCCATCCGGACAGTGTGCACCTCAAGACTAAGAAAGC	362
Db	364	IleAspCysThrTrpAsnIleGluValProAsnAsnGlnHisValLysValArgPheLys	383
Qy	363	ACTGTGCATCACTTGACCTGGAGCCTTCTGTGGAGCTGGCTGGCCGCTGCCT	422
Db	384	PhePheTyrLeuLeuGluProGly---ValProAlaGlyThrCysProLysAspTyr---	401
Qy	423	ACTCTGAAGTTCATGGCGCAGCAAG-----TGCTC	452
Db	402	---ValGluIleAsnGlyGlu-LysTyrCysGlyGluArgSerGlnPheValThrSe	420
Qy	453	CAACTCTGGG-----	462
Db	420	rAsnSerAsnLysIleThrVala:gpHeHisSerAspGlnSerTyrThrAspThrGlyPh	440
Qy	463	-----ATAGAGTGGACTCTCTCA-----	480
Db	440	eLeuAlaGluTyrLeuSerTyrAspSerSerAspProCysProGlyGlnPheThrCysAr	460
Qy	481	----GGTACCTGCATCAACCCCTTAACCTGGTGTGAGCGGTGTCACTCCCGCGCG	536
Db	460	gThrGlyArgCysIleA:glysGluLeuArgCysAspGlyTrpAlaAspCysThrAspHi	480
Qy	537	GGAGACAGAGATCGGTGTGTGCTCTCAGGACCAACTTCATCTTCAGGTGTACTC	596
Db	480	sSerAspGluLeuAsnGlySerCysAspAlaGlyHisGlnPheThrCysLysAsnLysPh	500
Qy	597	ATCTCAGAGGAAGTCTCTGGCACCTGTGTGCAAA-----GACGACTGGAAACGAGACTA	650
Db	500	eCysLysProLeuPheTrp-----ValCysAspSerValAsnAspCysGlyAspAsnSe	518
Qy	651	CGGGCGGGCGCTCGAGGACATGGGCTATAGATATATTTTACTCTAGCCAGGA--	708
Db	518	raspGluGlnGlyCysSerCysProAla-----GlnThrPheArgCysSerAsnGlyLy	536
Qy	709	-----ATAGTGGATGACAGCGGATC	728
Db	536	sCysLeuSerLysSerGlnGlnCysAsnGlyLysAspAspCysGlyAspGlySerAspGl	556
Qy	729	CACGAGCTTATGAACCTGAACACA-----	753

Search completed: July 6, 2004, 15:04:07
 Job time : 48 secs

Db	556	uAlaSerCysProLysValAsnValValThrCysThrLysHisThrTyrArgCysLeuAs	576
Qy	754	-----AGTCCGGCAATTCGATATCTATAAAAACTGTACACAGTATGTC	800
Db	576	nGlyLeuCysLeuSerLysAsnProGluCysAspGlyLysGluAspCysSerAspGl	596
Qy	801	CTGTCTCTCAAAAGCAGTGGTTCTTTACGCTGTATAGCTCGGGGTCAACTTGAATC	860
Db	596	ySerAspGluLys-----AspCysAspCysGlyLeuArgSe	608
Qy	861	A---AGCCGCCAGAGAGGATTTGGCGCGAGAGCGGCTCCCGGGGCCCTGGCCCTG	917
Db	608	rPheThrArgGlnAlaArgValValGlyGlyThrAspAlaAspGluGlyLysProTr	628
Qy	918	GCAGSTCAGCTGCAC---GTCCAGAACGTCACGCTGGAGGCTCCATCATCACC	974
Db	628	pGlnValSerLeuHisAlaLeuGlyGlnGlyHisIleCysGlyAlaSerLeuLeuSerPr	648
Qy	975	CGAGTGGATCGTGACAGCGCCCACTGCTGTGAAAAACCT-----CTTAACAA	1022
Db	648	oAsnTrpLeuValSerAlaAlaHisCysTyrIleAspAspArgGlyPheArgTyrSerAs	668
Qy	1023	TCCATGCGCATTTGACGGCATTTCCGGGATTTTGAGACATCTTCTGTATGTGAGC	1082
Db	668	pProThrGlnTrpThrAlaPheLeuGlyLeuHisAspGlnSer---GlnArgSerAlaPr	687
Qy	1083	CGGATACCAA-----GTAGAAAAGTGATTTCTCATCCAAATATATGACTCCAGAC	1133
Db	687	oGlyValGlnGluArgLeuLysArgIleIleSerHisProPheAsnAspPheTh	707
Qy	1134	CAGAACATGACATTTGGCTGTATGAGCTGCAGAGCCTCTGACTTTCAACACCTAGT	1193
Db	707	rPheAspTyrAspIleAlaLeuLeuGluLeuGluLysProAlaGluTyrSerSerMetVa	727
Qy	1194	GAACACAGTGTCTGCCCAACCCAGCATGTGTGCAGCCAGACAGCTCTGCTGGAT	1253
Db	727	lArgProIleCysLeuProAspAlaSerHisValPheProAlaGlyLysAlaIleTrpVa	747
Qy	1254	TTCCGGGTGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTGTGAACCTGCGCAA	1313
Db	747	lThrGlyTrpGlyHisThrGlnTyrGlyGlyThrGlyAlaLeuIleLeuGlnLysGlyGl	767
Qy	1314	GGTGCTTCTCATTGACACACAGAGATGCAACACAGATATGTCTATGACAACTGATCAC	1373
Db	767	uIleArgValIleAsnGlnThrThrCysGluAsn-----LeuLeuProGlnGlnIleTh	785
Qy	1374	ACCAGCCATGATCTGTCCCGCTTCTCTGCAGGGGAAAGCTCGATTTCTTTCAGGGTGCACAG	1433
Db	785	rProArgMetMetCysValGlyPheLeuSerGlyGlyValAspSerCysGlnGlyAspSe	805
Qy	1434	TGAGGGGCTCTG-----GTCACTTCGAAGAACAATATCTGTGTGGCTGTAGGGGATAC	1487
Db	805	rGlyGlyProLeuSerSerValGluAlaAspGlyArgIle---PheGlnAlaGlyValVa	824
Qy	1488	AAGCTGGGTCTGTGGCTGTGCCAAAGCTTACAGACCAGAGAGTGTACGGGAATGTGATGT	1547
Db	824	lSerTrpGlyAspGlyCysAlaGlnArgAsnLysProGlyValTyrThrArgLeuProLe	844
Qy	1548	ATTCACGACTGGATT	1563
Db	844	uPheArgAspTrpIle	849

DR WPI; 2000-116363/10.
 DX N-PSDB; AAZ29636.

XX Novel cell surface antigen useful to treat colon and prostate cancer.

XX Claim 1; Fig 1; 58pp; English.

XX The present sequence is the 20p1f12 protein (also known as the TMPRSS2 protein) which is a prostate-specific, androgen-regulated, cell surface serine protease. It is a glycosylated type II transmembrane protein with an extracellular C-terminal serine protease domain, a scavenger receptor cysteine-rich domain, an LDL receptor class A domain and a predicted transmembrane domain. Host cells can be transformed to produce this protein, using vector containing 20p1f12/TMPRSS2 gene (also designated 20p1f12-GTC1, as deposited with ATCC accession number 207097). Anti-20p1f12/TMPRSS2 antibodies may be used as therapeutic agent for prostate and colon cancers, to image prostate cancer cells and prostate tumours, to identify ligands and cellular constituents that bind to a 20p1f12/TMPRSS2 gene product and for use as cancer vaccines

XX Sequence 492 AA;

Alignment Scores:

2red. No.: 7,27e-222 Length: 492
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.47% Indels: 0
 DB: 3 Gaps: 0

US-09-323-597C-1 (1-1738) x AAY44406 (1-492)

QY 112 ATGGCTTTGAACACTAGGCTACCCAGCTATTGACCTTCTATGAAACCATGATAC 171
 DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAACCCCTATCCCGCAGCCACTGTGGTCCCACTGTCTAGAGTGCAT 231
 DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 QY 232 CCGGCTCAGTACTACCGCTCCCGCTGCCAGTACGCCCGAGGGCTCTCAGCAGGCT 291
 DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCGCTGTGACGAGCCCAATCCCATCCGAGCAGCTGTGCACTCAAG 351
 DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGACCTTCTCTGGAGCTGGCTG 411
 DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 QY 412 GCGGTGGCCTACTCTGGAAGTTCTGGGAGCAAGTCTCTCACTCTGGGATAGATGC 471
 DB 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyLeuGluCys 120
 QY 472 GACTCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGATGGCTGTCACTGGCCCC 531
 DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGGAGGAGGAGATGGTGTTCGCTCTACGAGCCAAACTTCATCTTCAGGTG 591
 DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGAGAGTCTGGCACCTGTGTGCAAGACGACTGGAAACGAGACTAC 651
 DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GCGGGCGGCTGTCAGGAGACATGGGCTATAGATATATTTTACTTAGCAAGGATA 711
 DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATGACGCGGATCCACGAGCTTTATGAACCTGAACACAGTCCGCGCAATGTCAT 771

DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAMAACTGTACCAAGTGTCTCTTCAAAAGCAGTGGTCTTCTTACGC 831
 DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTTGGCGGTCAACTTGAACCTCAAGCCGCGCAGACGAGGATTTGGGCGCGGAG 891
 DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGGCTCCCGGGGCGCTGGCCCTGGCAGTCTGAGCTGCACGCTCCAGACGCTCCAGTG 951
 DB 261 SerAlaLeuProGlyAlaTrpTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TSCGGAGGCTCCATCATCACCCCGAGTGTGATCGTGCAGCCGCCCTCCGCTCGGAAAAA 1011
 DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTAACAATCCATGGCATTCGAGCGGCAATTTGGGGATTTTGAGACAATCTTCATG 1071
 DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGGAGCCGCGATACCAAGTAGAAAAAGTCAATTTCTCATCAAAATTTAGTCTCCAG 1131
 DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 QY 1132 ACCAAGAACCAATCAGATTGCGTGTGATGAAGTCCAGAGGCTCTGACTTTCAACAGCTTA 1191
 DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 QY 1192 GTCAACACCGAGTGTCTGCCCAACCCAGGCATCATCTGCAGCCAGAACAGCTCTGCTGG 1251
 DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 QY 1252 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGGAAGACCTTCAGAGTGTCTGAACGCTGC 1311
 DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 QY 1312 AAGTGCTTCTCATTTGAGACACAGATGCAACAGAGATATCTATATGCAACCTGATC 1371
 DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 QY 1372 ACACCGACCATGATGTGTCCGGCTTCTTCAGGGGGAACCTCGAATTTCTCCAGGCTGAC 1431
 DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTGGAGGCGCTCTGTGCTCACTTCGAAGAACAAATATCTGTGGCTGATAGGGGTACAGC 1491
 DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 QY 1492 TGGGGTTCTGGCTGTGCCAAAGCTTACAGACGAGGAGTGTACGGGAATGTGATGGTATTC 1551
 DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 QY 1552 ACAGACTCGGATTTTATCGACAAATGAGCGGCAGACGCG 1587
 DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2

AAAY77726
 ID AAAY77726 standard; protein; 492 AA.

XX AC AAAY77726;

XX 12-MAY-2000 (first entry)

XX Human tumour suppressor TMPRSS2 polypeptide.

XX Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
 KW gene therapy; protein therapy.

XX Homo sapiens.

XX W020000605-A1.
 XX PD 06-JAN-2000.
 XX PF 29-JUN-1999; 99WO-US014622.
 XX PR 29-JUN-1998; 98US-0091044P.
 XX (MYRI-) MYRIAD GENETICS INC.
 XX PI Wong AKC, Tavtigian SV, Teng DHF;
 XX DR WPI; 2000-170914/15.
 XX DR N-PSDB; AA287786.
 XX PT Novel tumor suppressor TMRPS2 used for the diagnosis and prognosis of
 XX PT human cancer.
 XX PS Claim 55; Page 77-79; 89pp; English.
 XX CC The invention provides a new tumor suppressor gene, designated TMRPS2.
 CC The TMRPS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and prognosing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with anti
 CC -cancer or therapeutic properties. The polypeptides are also useful for
 CC rational drug design. The TMRPS2 polynucleotides and polypeptides may be
 CC used for gene therapy and protein therapy. The present sequence
 CC represents the TMRPS2 polypeptide
 XX
 XX SQ Sequence 492 AA;

Alignment Scores:
 Pred. No.: 7,27e-222 Length: 492
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.47% Indels: 0
 DB: 3 Gaps: 0

US-09-323-597C-1 (1-1738) x AAV77726 (1-492)

QY	112	ATGGCTTTGAATCAGGCTGACACCCAGCTATTGGACCTTACTATGAAACCATGATAC	171
Db	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
QY	172	CAACCGGAAACCCCTATCCCGACAGCCCACTGFGTCCCACTGCTCTACGAGGTGCAT	231
Db	21	GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis	40
QY	232	CCGGCTCAGTACACCGCTCCCGCTGCCAGTACGCCCGGAGGGTCTTCGACGAGGCT	291
Db	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
QY	292	TCCAAACCCCTGCTGTCAGCGAGCCAAATCCCATCCCGGACAGTGCACCTCAAAG	351
Db	61	SerAsnProValValCysThrGlnProIlySerProSerGlyThrValCysThrSerLys	80
QY	352	ACTAAGAAAGCACTGTCATCACCTGCACCTCGGACCTTCTCTCGGAGAGTGGCTG	411
Db	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu	100
QY	412	CGCGCTGGCTACTCTGAAATTCATCGGCGAGCAAGTCTCCAACTCGGAGTAGAGTGC	471
Db	101	AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
QY	472	GACTCTTCAGGTACTCTCATCAACCCCTCTAACTGGTGTGATGGCGTGTACACTGCCCC	531
Db	121	AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro	140
QY	532	GGCGGGAGGACGGAATCGGTGTTGCTTACGGACCAACTTCATCTTCAGTTCAGT	591
Db	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160

RESULT 3

AA218096 standard; protein; 492 AA.

ID

XX

QY	592	TACTCATCTCAGAGGAAGTCTCTGGCAACCTGTGTGCCAGACGACTGGACGAGACTAC	651
Db	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr	180
QY	652	GGGCGGGGGCTGTCAGGGACATGGCTATAAGAAATAATTTTACTCTAGCCAAAGATA	711
Db	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200
QY	712	GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAGTCCGGGCAATCTCGAT	771
Db	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
QY	772	ATCTATAAAAACTGTACACAGTGCCTCTTCTCAAAAGCAGTGTCTTCTTTTACGC	831
Db	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
QY	832	TCTATAGCTGGGGGTCAACTGAACCTCAAGCCCGCAGAGCAGGAATGTGGCGCGAG	891
Db	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu	260
QY	892	ACGGCGCTCCCGGGGGCTGGCCCTGGCAGGTGCAGCTGCAGTCCAGAACCTCCACGTG	951
Db	261	SerAlaLeuProGlyAlaIleTrpTrpGlnValSerLeuHisValGlnAsnValHisVal	280
QY	952	TCCGAGGCTCCATCATCACCCCCGAGTGGATCGTGACAGCCCGCCACTGCGTGAAAAA	1011
Db	281	CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys	300
QY	1012	CCTCTTACATCCATGGCATGGACCGCATTTGGGGGATTTTGAGACAACTCTTTCATG	1071
Db	301	ProLeuAsnAsnProTrpHisIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
QY	1072	TTCTATGGAGCCGGATACCAAGTAGAAAAGTATTCTTCATCCAAATTCATCTCAAG	1131
Db	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys	340
QY	1132	ACCAAGAACATGCAATTGCGCTGTAGAGCTGCGAGAACCTCTGACTTTCACGACCTA	1191
Db	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysThrSerGluValLeuAsnAlaAla	360
QY	1192	GTGAACACCATGTGTGTGCCCAACCCAGGCATGTGTGTCAGCCAGAACACAGCTCTCTG	1251
Db	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr	380
QY	1252	ATTTCGGGTGGGGGGCCACCAGAGAGAAAGGAGAACCTCAGAGTGTCTGACCGTGC	1311
Db	381	IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
QY	1312	AAAGTGTCTTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC	1371
Db	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle	420
QY	1372	ACACCAAGCCATGATCTGTCCCGCTTCTCGAGGGGAAAGTGTGATCTTTCGACGGGTGAC	1431
Db	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
QY	1432	ACGTGAGGGCTCTGTCATCTTCGAGAACATATCTGTGTGCTGATGCGGATACAAGC	1491
Db	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIlePheulleIleAspThrSer	460
QY	1492	TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTC	1551
Db	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
QY	1552	ACGACTGTGATTTATCGAACAAATGAGGCGCAGCGGC	1587
Db	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly	492

AAE18096;
 07-MAY-2002 (first entry)
 Human 20P1F12-GTCl protein.
 Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCl; cell growth; neoplasm;
 cancer; vaccine; human.
 Homo sapiens.
 WO200204953-A2.
 17-JAN-2002.
 12-JUL-2001; 2001WO-US022168.
 12-JUL-2000; 2000US-00615295.
 (AGEN-) AGENSYS INC.
 Safertan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
 Challita-Bid EW;
 WPI; 2002-154967/20.
 N-PSDB; AAD28778.
 Examining a biological sample for evidence of dysregulated cellular
 growth, comprises comparing the status of prostate-specific, androgen-
 regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
 normal sample.
 Claim 40; Fig 1; 161pp; English.
 The present invention relates to methods and compositions for the
 diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
 kidney cancer derived from or based on a normally prostate-specific,
 androgen regulated, cell membrane associated secreted serine protease
 termed 20P1F12/TMPRSS2. The invention further relates to a method of
 examining a biological sample for evidence of dysregulated cellular
 growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
 designated 20P1F12-GTCl) in the sample to the status of 20P1F12/TMPRSS2
 in a corresponding normal sample. The invention also relates to 20P1F12/
 TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
 invention are used for examining a sample such as blood, serum, stool,
 urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
 The dysregulated cell growth is indicative of bladder cancer, lung
 cancer, kidney cancer or ovarian cancer. It is useful for identifying
 evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
 portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
 expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
 lung cancer, ovarian cancer or metastatic cancer. The present sequence is
 human 20P1F12-GTCl protein

Alignment Scores:
 Pred. No.: 7,278-222 Length: 492
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.47% Indels: 0
 DB: 5 Gaps: 0

US-09-323-597C-1 (1-1738) x AAE18096 (1-492)

QY 112 ATGGCTTGAAGTCTAGGCTCACCAGCTATTGACCTTACTATGAACCAATGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAAAACCCCTATCCCGCACAGCCCACTGTGTCCCGACCTGTCTACAGATGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40

QY 232 CGGGCTCAGTACTACCCGTCCTCCCGTGGCCAGTACGCCCGAGGGTCTCTGACGAGGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCGCTGCTCTGACGAGCCCAATCCCATCCGAGACATGTGCACCTCAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAAGCACTGTGCATCACCTTGACCTCGGGGACCTTCCTCGTGGAGCTGGCGTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 GCGCTGGCTACTCTGGAAGTTCATGGGAGCAAGTCTCCAACTCTGGATAGAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCTCAGGTACCTGCATCAACCCCTCTAATCTGGTGTGATGGCTGTGCACACTGCCCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGGGAGACGAGAAATCGTGTGCTTCACGACCCAACTTCATCTCTTCAGGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGAGGAGTCTTGGCACCTGTGTCACAGACGACCTGGACGAGAACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GCGCGGCGGCTCTGACGAGGACATGGGCTATAAGAATAATTTTACTTACCCAGGAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATCAGCGGATCCACAGCTTTATGAACCTGAACACAGTCCCGCAATGCGAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACACACAGTGTGCTGTCTTCTTCAAAAGCAGTGGTTCCTTACGC 831
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTTGGGGGTCACTTGAACCTCAGCCGCGCAGGACGAGTATGTGGCGGGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGTGTGACCTGCAGCTCCAGAACGTCCACGTG 951
 Db 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TGGGAGGCTCCATCATCACCCTCCGAGTGATCGTGACAGCCGCCCTCCGCTGGAAAAA 1011
 Db 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTAACAATCCATGGCATTGCGGGGATTTGGAGACAATCTTTTCATG 1071
 Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGGAGCCGGATACCAAGTAGAAAAGTAGTTTCTCATCAAATTATGACTCCAAG 1131
 Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 QY 1132 ACCAAGAACATACATTTGGCTGTGATGACAGCTCCAGAGGCTCTGACTTTTCAACGACCTA 1191
 Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 QY 1192 GTGAACCAAGTGTCTTGGCCCAACCCAGGATGATGTGACGACGAGCAACAGCTCTGCTGG 1251
 Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 QY 1252 ATTTCCGGTGGGGGCGCCACCGAGGAGAAAGGAAGCACTCAGAGTGTGAGAGCTGCC 1311
 Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyGlyThrSerGluValLeuAsnAlaAla 400

QY 1312 AAGGTGCTTCTATTGAGACACAGAGATGCAACAGCAGATATCTCTATGACAACTGATC 1371
 Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 QY 1372 ACACCGCCATGATCTGTGGCGCTCTCCAGCGGAGCTCGATTCTTCCAGGGTGAC 1431
 Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTGGAGGGCTCTGTGCTCACTTCGAAGAACAAATATCTGTGGCTGATAGGGGATACAAGC 1491
 Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpLrpLeuIleGlyAspThrSer 460
 QY 1492 TGGGTCTGCTGTGTCACAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATTC 1551
 Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 QY 1552 ACGGACTGGATTTATCGACAAATGAGGCGCAGACGGC 1587
 Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 4

AAE18098
 ID AAE18098 standard; protein; 492 AA.

AC AAE18098;

XX 07-MAY-2002 (first entry)

XX Human 20P1F12/TMPRSS2 mutant protein (S441A).

XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
 cancer; vaccine; human.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 441 /note= "Wild type Ser substituted with Ala"

XX W0200204953-A2.

XX 17-JAN-2002.

XX 12-JUL-2001; 2001WO-US022168.

XX 12-JUL-2000; 2000US-00615285.

XX (AGEN-) AGENSYS INC.

XX Saferran D; Raitano AB, Hubert RS, Jakobovits A, Paris M;
 Challita-Bid PM;

DR WPI; 2002-154967/20.

XX Examining a biological sample for evidence of dysregulated cellular
 growth, comprises comparing the status of prostate-specific, androgen-
 regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
 normal sample.

XX Example 10; Page; 161pp; English.

XX The present invention relates to methods and compositions for the
 diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
 kidney cancer derived from or based on a normally prostate-specific,
 androgen regulated, cell membrane associated secreted serine protease
 termed 20P1F12/TMPRSS2. The invention further relates to a method of
 examining a biological sample for evidence of dysregulated cellular
 growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
 designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TMPRSS2
 in a corresponding normal sample. The invention also relates to 20P1F12/
 TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
 invention are used for examining a sample such as blood, serum, stool,

CC urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
 CC The dysregulated cell growth is indicative of bladder cancer, lung
 cancer, kidney cancer or ovarian cancer. It is useful for identifying
 evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
 portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
 expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
 lung cancer, ovarian cancer or metastatic cancer. The present sequence is
 human 20P1F12/TMPRSS2 mutant protein (S441A). Note: This sequence is not
 shown in the specification but is derived from wild type 20P1F12/TMPRSS2
 CC protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18098)

XX Sequence 492 AA;

Alignment Scores:
 Pred. No.: 1,31e-221 Length: 492
 Score: 2714.00 Matches: 491
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.80% Mismatches: 0
 Query Match: 85.37% Indels: 0
 DB: 5 Gaps: 0

US-09-323-597C-1 (1-1738) x AAE18098 (1-492)

QY 112 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCATGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAAACCCCTATCCCGCAGCCACCTGTGTCGCCACTCTCTACGAGGTGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 QY 232 CCGGCTCAGTACTACCCGTCCTCCCGTCCCGCAGTACGCCCGAGGTCCTGACGCGGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAACCCGTCGTCTGCGACGAGCCCAATCCCATCCGCGGACAGTGTGCCTCAAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCTCTGGGACCTTCTCTGTGGAGCTGCGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 GCGCTGGCTACTCTGGAAAGTTTCATGGGAGCAAGTCTCCACTCTGGATAGATGC 471
 Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCCTCAGTACCTGCATCAACCCCTCTAACTGTGTGATGGCGTGTCACTGCGCTGCC 531
 Db 121 AppSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGCGGAGGACGAGAAATCGTGTGTTCGCTCTACGACCCAACTTCATCTCAGGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGGAGTCTCGGACCCCTGTGTGCAAGACGACCTGGAACGAACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GCGCGGCGGCTCTGACGAGGACATGGCTATAAGATAATTTTCTCTCTACCCAGGAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATGACGCGGATCCACGACCTTTATGAACCTGAACACAGTCCCGCAATGTCAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACACAGTGTGCTGTTCTTCAAAACAGCTGGTTCCTTACGC 831
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGATATAGCTTGGCGGTCAACTTGAACCTCAGCCCGCAGAGGAGTTGTGGCGGCGAG 891

Db 241 CysIleAlaCysGlyValAlaAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Qy 892 AGCGCGTCCCGGGGGCTGGCCCTGGCAGGTGACCTGACGCTCCAGACGTCCAGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnValHisVal 280
Qy 952 TGGGAGGCTCCATCATCACCCTCCGAGTGGATCTGTACAGCGCCCTGCTGGTGA 1011
Db 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
Qy 1012 CCTCTTAAATCCATGTCATTCGACGCGCATTTGGGGGATTTTGAGACAACTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrPheTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Qy 1072 TTTATGAGCGGATACCAAGTATGAAAAAGTATTTCTATCCAAATATGATCCAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Qy 1132 ACCAAGAACATGACATTCGCTGATGAGCTGACAGAGCTCTGACTTTCAACGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Qy 1192 GTGAACACGATGTGTCTCCCAACCCAGGCGATGATGCTGCAGCCAGACACACTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Qy 1252 ATTTCCGGTGGGGGGCCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Qy 1312 AAGGTGCTCTTCATTTGACACACAGAGATGCAACAGCAGATATGCTATGACACCTGATC 1371
Db 401 LysValLeuLeuLeuLeuThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Qy 1372 ACACAGCAGCATGATCTGTGCGGGCTCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Qy 1432 AGTGGAGGCGCTCTGTCTGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1491
Db 441 AlaGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Qy 1492 TGGGTCTGCTGTGCGCAACCTTACAGACAGAGGTGTACGGGAATGTATGATATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Qy 1552 ACGGACTGATTTATCGCAAAATGAGGCGAGCGGC 1587
Db 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492

RESULT 5
AAE18100
ID AAE18100 standard; protein; 492 AA.
XX AC AAE18100;
XX DT 07-MAY-2002 (first entry)
XX DE Human 20P1F12/TMPRSS2 mutant protein (R252Q).
XX KW Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCL; cell growth; neoplasm;
XX KW Cancer; vaccine; human.
XX OS Homo sapiens.
XX OS Synthetic.
XX PH Key Location/Qualifiers
XX FT Misc-difference 252 /note= "wild type Arg substituted with Gln"
XX PN W0200204953-A2.
XX PD 17-JAN-2002.

XX 12-JUL-2001; 2001WO-US022168.
XX 12-JUL-2000; 2000US-00615285.
XX (AGEN-) AGENSYS INC.
XX Saferan D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
XX Challita-Bid PM;
XX WPI; 2002-154967/20.
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
XX normal sample.
XX Example 10; Page; 161pp; English.
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TMPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
XX designated 20P1F12-GTCL) in the sample to the status of 20P1F12/TMPRSS2
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12/TMPRSS2 mutant protein (R252Q). Note: This sequence is not
XX shown in the specification but is derived from wild type 20P1F12/TMPRSS2
XX protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX Sequence 492 AA;
XX Alignment Scores:
XX Pred. No.: 1.59e-221 Length: 492
XX Score: 2713.00 Matches: 491
XX Percent Similarity: 100.00% Conservative: 1
XX Best Local Similarity: 99.80% Mismatches: 0
XX Query Match: 85.34% Indels: 0
XX DB: 5 Gaps: 0
XX US-09-323-597C-1 (1-1738) x AAE18100 (1-492)
Qy 112 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Qy 172 CAACCGGAAACCCCTATCCCGACACGCCACCTGTGTGCTCCCACTGTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Qy 232 CCGCTCAGTACTACCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCC 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Qy 292 TCCAAACCCCGTGTCTCAGCGAGCCCAATCCCAATCCCAATCCCAATCCCAATCCCAATCC 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Qy 352 ACTAAGAAGACACTGTCATCACCTTCACCTGGGAGCTTCTCTCGTGGGAGCTGCCTG 411
Db 81 ThrLysLysAlaLeuLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100

Alignment Scores:

Pred. No.:	1,59e-221	length:	492
Score:	2713.00	Matches:	491
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.80%	Mismatches:	0
Query Match:	85.3%	Indels:	0
DB:	5	Gaps:	0

US-09-323-597C-1 (1-1738) x AAE18099 (1-492)

112	ATGGCTTTGAACTCAGGGTCACCAACGAGCTATTGGACTTACTATGAATAACCATGGATAC	171
1	MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
172	CAACCGGAAACCCCTATCCGGCACAGCCCACTGGTCCCACTGTCTACAGGTGCAT	231
21	GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
232	CGGGCTCAGTACTACCCGTCGCCGTCGCCAGHACGCCCGAGGGTCTCTGACGAGGCT	291
41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
292	TCCAAACCCGCTGCTCCAGCAGCCACAAATCCCACTCCGGGACAGTGTGCACCTCAAAG	351
61	SerAsnProValValCysThrGlnProIysSerProSerGlyThrValCysThrSerIys	80
352	ACTAAGAAAGCACTGTGATCACCTTGACCTCGGGACCTTCTCTCGTGGAGCTCGCGTG	411
81	ThrIysIysAlaIeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyValAlaLeu	100
412	CGCGTCGGCTACTCTCGAAGTTCATGGGACAGAGTCTCCAACTCTGGGATAGAGTGC	471
101	AlaAlaGlyLeuLeuTprLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
472	GACTCTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGCGGTGCACATGCCCC	531
121	AspSerSerGlyThrCysIleAsnProSerAsnTprCysAspGlyValSerHisCysPro	140
532	GGCGGGGAGCAGAGAATCGGTGTGCGCTCTACGGACCAAACTTCATCTCTCAGGTG	591
141	GlyGlyGluAspGluAsnArgCysValArgIeuTyrGlyProAsnPheIleLeuGlnVal	160
592	TACTCATCTCAGAGGAAGTCTCTGCACCTGTGTGCCAAGACGACTGGAAACGAACTAC	651
161	TyrSerSerGlnArgLysSerTprHisProValCysGlnAspAspTprAspGluAsnTyr	180
652	GGCGGGCGGCTCGACGGACATCGGCTATAAGATAATTTTACTCTTAGCCRAGGATA	711
181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200
712	GTGGATGCACAGCGGATCCACAGCTTTATGAACCTGAACCAAGTCGCGGCAATGTGCT	771
201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
772	ATCTATATAAAACGTGTACCAACAGTAGTCCCTGTCTTCAAAAGCAGTGGTTCCTTACG	831
221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuGln	240
832	TGTATAGCTCGCGGGTCAACTTCAACTCAAGCGCCGACAGACAGGATGTGGCGGGCAG	891
241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu	260
892	AGCGGCTCTCCGGGGGCTGGCCTGCGAGGTGAGCTGCGACGTCCAGACGCTCCAGGTG	951
261	SerAlaLeuProGlyAlaTprProTprGlnValSerLeuHisValGlnAsnValHisVal	280
952	TGGGGAGGCTCCATCATCCCCCGAGTGGATCGTAGACGCGGCCCACTGCTGGAAAAA	1011
281	CysGlyGlySerIleIleThrProGluTprIleValThrAlaAlaHisCysValGluLys	300
1012	CTCTTTAAACAATCCATGGCATTTGACCGGCAATTTTGGAGCAATCTTTTCATG	1071
301	ProLeuAsnAsnProTprHisTprThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320

QY	1072	TTCTATCGAGCCGGATACCAAGTAGAAAAAGTCATTTCTCATCCCAATATGACTCCCAAG	1133
Db	321	PhetyrGlyAlaGlytyrGlnValGluLeuValIseHisProAntytrAspSerLys	340
QY	1132	ACCAAGAACAAATGACATTGGCGCTGATGAAGCTGCAGAAGCCTCTGACTTTTCAAACGACCTA	1191
Db	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
QY	1192	GTGAACACCACTGTGCTGCCCAACCCAGGCGATCATCTGCAGCCAGACAGCTCTGCTGG	1251
Db	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuGlyTyr	380
QY	1252	ATTTCCGGGTGGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAGTGTCTGAACGCTGCC	1311
Db	381	IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
QY	1312	ARGTGTCTTCTCATTTGAGACACACAGAGATCAACACAGACAGATATCTCTATGCACACCTGATC	1371
Db	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgtyrValLysAspAsnLeuIle	420
QY	1372	ACACCAACCCATGATGTGTGCGCGCTTCCTGTCAGGGGAAACGTCGATTTTTCGCCAGGGTGAC	1431
Db	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
QY	1432	ACTGGAGCGCTCTGCTCACTTCGAGAGACAAATATCTGTGGCTGATAGGGGATACAGC	1491
Db	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrPheLeuIleGlyAspThrSer	460
QY	1492	TGGGGTTCTGGCTGTGCCAAAGCTTACACACCAAGGAGTGTACGGGAATGTGATGGTATTC	1551
Db	461	TyrGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
QY	1552	ACGAGCTGGATTATTCGACAAATGAGCGGCAGACGGC	1587
Db	481	ThrAspTyrPheTyrArgGlnMetArgAlaAspGly	492
RESULT	7		
ID	AAE18101		
XX	AAE18101	standard; protein; 492 AA.	
AC	AAE18101;		
XX			
DT	07-MAY-2002	(first entry)	
XX			
DE	Human 20P1F12/TMPRSS2	mutant protein (R255Q).	
XX			
KW	Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCl;	cell growth; neoplasm;	
KW	cancer; vaccine; human.		
XX			
OS	Homo sapiens.		
XX	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference 255	/note= "Wild type Arg substituted with Gln"	
XX			
XX	WO200204953-A2.		
XX			
PD	17-JAN-2002.		
XX			
XX	12-JUL-2001; 2001WO-US022168.		
PF			
XX			
PR	12-JUL-2000; 2000US-00615285.		
XX			
PA	(AGEN-) AGENSYS INC.		
XX			
PI	Saferan D, Raitano AB, Hubert RS, Jakobovits A, Paris M;		
PI	Challita-Sid PM;		
XX			
DR	WPI; 2002-154967/20.		
XX			
FT	Examining a biological sample for evidence of dysregulated cellular		

PT growth, comprises comparing the status of prostate-specific, androgen-regulated, secreted serine protease, 20P1F12/TPRSS2, in a corresponding normal sample.

XX Example 10; Page; 161pp; English.

XX The present invention relates to methods and compositions for the diagnosis and therapy of prostate, colon, bladder, lung, ovarian and kidney cancer derived from or based on a normally prostate-specific, androgen regulated, cell membrane associated serine protease termed 20P1F12/TPRSS2. The invention further relates to a method of examining a biological sample for evidence of dysregulated cellular growth comprises comparing the status of 20P1F12/TPRSS2 gene (also designated 20P1F12-GT1) in the sample to the status of 20P1F12/TPRSS2 in a corresponding normal sample. The invention also relates to 20P1F12/TPRSS2 polynucleotides and their corresponding proteins. Methods of the invention are used for examining a sample such as blood, serum, stool, urine, semen, or biopsy tissue for evidence of dysregulated cell growth. The dysregulated cell growth is indicative of bladder cancer, lung cancer, kidney cancer or ovarian cancer. It is useful for identifying evidence of a neoplasm in a sample. Vaccines comprising an immunogenic portion of 20P1F12/TPRSS2 are useful for inhibiting growth of a cell expressing 20P1F12/TPRSS2 in a patient suffering from bladder cancer, lung cancer, ovarian cancer or metastatic cancer. The present sequence is human 20P1F12/TPRSS2 mutant protein (R255Q). Note: This sequence is not shown in the specification but is derived from wild type 20P1F12/TPRSS2 protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)

XX Sequence 492 AA;

Alignment Scores:

Pred. No.: 1.59e-221 Length: 492
Score: 2713.00 Matches: 491
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.90% Mismatches: 0
Query Match: 85.34% Indels: 0
DB: Gaps: 0

US-09-323-597C-1 (1-1738) x AAE18101 (1-492)

QY 112 ATGGCTTTGAAGTCACTGAGGTCACACAGCTATTGGACTTACTATGAAACACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCGACACACCCCACTGTGTGTCCTCCCACTGTCTACGAGGTCCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCGTCCTCCCGTCCCGCTCCAGTACGCCCCGGGTCTTGACCGAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCACACCCCTGCTGTCACGCGAGCCCAATCCCGATCCGCGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAACACCTGTCATCCTGACCTGACCTGGGACCTTCTCTGTTGGAGCTGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTACTCTGGAAGTTTCATGGGACAGAGTGTCCCACTTCGGATAGAGTGC 471
DB 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGTACTGTCATCAACCCCTCTAATGGTGTGATGGCTGTGTCACACTCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GGGGGGAGGACGAGAAATCGGTGTGTTGCGCTCTACGACCAAACTTCATCTTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAGTCTCTGGCACCCCTGTGTGCCAGACGACTGGAGACGAGACTAC 651

DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
QY 652 GGGCGGGCGGCTGTCAGGAGCATGGGCTATATAGATAATATTTTACTCTACCCAGGAATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAAACTGAAACACACAGTCCCGCAATGTGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTACACAGTGTGCTCTCTCAAAAGCAGTGTGTTCTTTACGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCCGACAGCAGGATTTGGCGCGCGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerGlnIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGCGCTGCGCCCTGGCAGGTGAGCTCCACGTCCAGAACGTCACGTG 951
DB 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCATCACCCCGAGTGTGATGTCAGACCGCCGCTCGCTGGGAAA 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CTTCTTAACAACTTCATGAGTTCGAGCGCATTTGGCGGATTTTGGAGACAATCTTTCATG 1071
DB 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGAGCGGATACCAAGTAGAAAAAGTGAATTTCTATCCAAATATGATGCTCAAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTCGCTGATGAAGCTGCAGAGCCCTCTGACTTTTCAACGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCAAGTGTGTGCTGCCAACCCAGCATGATGTCGACCCAGAACAGCTCTGCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGCCACCCAGGAGAAAGGAGACCTCAGAGTGTCTGAACGCTGCC 1311
DB 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGCTCTATGACAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyr-ValTyrAspAsnLeuIle 420
QY 1372 ACACGACCATCATCTGTGCGGCTTCTGCAAGGGAAGCTCGATTTCTGCCAGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCTCTGTCTCACTTCGAAAGAACATATCTGTGCTGATAGGGATACAAAG 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGGCTGTGCCAAAGCTTACAGACGAGAGTGTACGGAGTGTGATGTGTTATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGATTTATCGACAAATGAGGCGCAGCGC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 8

ADB75591

ID ADB75591 standard; protein; 492 AA.

XX AC ADB75591;

XX


```

DT 04-DEC-2003 (first entry)
XX XX
DE Prostate cancer marker protein.
XX XX
XX Prostate; cancer; cytostatic; gene therapy; marker.
XX XX
OS Homo sapiens.
XX XX
XX WC2003009814-A2.
XX XX
XX 06-FEB-2003.
XX XX
XX 25-JUL-2002; 2002WO-US023913.
XX XX
XX 25-JUL-2001; 2001US-0307982P.
XX PR 22-AUG-2001; 2001US-0314356P.
XX PR 25-SEP-2001; 2001US-0325020P.
XX PR 12-DEC-2001; 2001US-0341746P.
XX PR 05-MAR-2002; 2002US-0362158P.
XX XX
XX (MILL-) MILLENNIUM PHARM INC.
XX PA
XX Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;
XX PI Hoerssh S, Kamatkar S, Wonsey AM, Glatt K, Zhao X, Anderson D;
XX XX
XX WPI; 2003-248033/24.
XX XX
XX New nucleic acid molecule, useful for diagnosing or treating prostate
XX PT cancer.
XX PT
XX PS Disclosure; SEQ ID NO 415; 99pp; English.
XX XX
XX The invention relates to newly discovered cancer markers associated with
XX CC the cancerous state of prostate cells. Also disclosed is a method of
XX CC assessing whether a patient is afflicted with prostate cancer. The method
XX CC of the invention involves assessing whether a patient is afflicted with
XX CC prostate cancer by comparing the level of expression of a marker in a
XX CC patient sample and the normal level of expression of the marker in a
XX CC control non-prostate cancer sample, where a significant increase in the
XX CC level of expression of the marker in the patient sample and the normal
XX CC level indicates that the patient is afflicted with prostate cancer.
XX CC Nucleic acids of the invention are useful for diagnosing or treating
XX CC prostate cancer, and may be useful in gene therapy. Sequences given in
XX CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 492 AA;

Alignment Scores:
Pred. No.: 2,35e-221 Length: 492
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.53% Mismatches: 0
Query Match: 85.28% Indels: 0
DB: 7 Gaps: 0

US-09-323-597C-1 (1-1738) x ADB75591 (1-492)
QY 112 ATGGCTTTGAACTCAGGTCAACACAGCTATTGACCTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGGCCACTGTGGTCCCACTGTCTACGAGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTTCACTACTACCGTCCCGCTGCGCCAGTACGGCCGCGAGGCTCTGACGAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTCGACGAGGCCCAATCCCATCCCGGACAGCTGTGCACTCAAG 351

```

```

61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
352 ACTAAGAAAGCACTGTGCATCACCCTTGACCTCGGGAGCTTCTCTGCTGGGAGCTCGCTG 411
81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
412 GCCCTCGCCTACTCTGAAAGTTCAATGGGAGCAAGTGTCTCAACTCTGGGATAGAGTGC 471
101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
472 GACTCTCAGGTACTGCATCAACCCCTCTAACTGGTGTGATGGCTGTGTCACTGCCCC 531
121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
532 GCGCGGAGGACGAGAATCGGTGTGCTCTACGACCAAACTTCTCATCTTCAGGTG 591
141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
592 TACTCATCTCAGAGAAAGTCTTGGCACCTGTGTGCCAAGACGACGACGAGAACTAC 651
161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
652 GGGGGGGCGCTCGACGACATGGCTATAAGAAATATTTTACTCTAGCAAGGAATA 711
181 GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
712 GTGATGACAGCGGATCCACAGCTTTATGAACATGAACACAACTGCGGCAATGTGAT 771
201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
772 ATCTATAAAACCTGTACCAAGTATGCTGTCTTCAAAAGCAGTGTCTTCTTCACG 832
221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
832 TGTATAGCCTCGCGGGTCAACTTGAACCTCAAGCCCGCAGACGAGATGTGGCGGCGAG 891
241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
892 AGCGGCTCCGGGGGCTGGCTGCGAGTCAAGCTGACGTCAGCTCCAGACAGTCCAGTG 951
261 SerAlaLeuProGlyAlaIleProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
952 TGCAGAGCTCCATCATCACCCCCGAGTGGATCGTGACAGCCGCCCACTCGTGGAAAAA 1011
281 CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
1012 CTCTTAACAATCCATGGCATTTGACCGGCAATTCGCGGGATTTTGACACAACTTTTATG 1071
301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
1072 TTCTATGAGCCGGATACCAAGTAGAAAAAGTGTCTCTCATCCAAATTTACTCTCAAG 1131
321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
1132 ACCAAGAACAAATGACATTTGCGTGTGATGAAGCTGACAGAGCTCTGTGACTTTCAAGACCTA 1191
341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
1192 GTGAACCCAGTGTCTGCCCCAACCCAGGCAATGATGCTGCGACGACAGCTCTGCTGG 1251
361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
1252 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAAAGACCTCAGAAAGTGTGAAGCTGCC 1311
381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
1312 AAGGTCTCTCTCATTTGACACAGAGTGCACAGCAGATATGCTATGATGACACCTGATC 1371
401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
1372 ACACCCGCCATGATCTGTGCGCGCTTCTGACGGGAAACGTCGATTTCTTCCACGGGTGAC 1431

```

Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTGGAGGCGCTCTGCTCACTTCGAGCAACAATATCTGGTGGCTGATAGGGGATCAAGC 1491
 Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 QY 1492 TGGGGTTCTGGCTGTGCAAGCTTACAGACCAAGGAGTGTACGGGATGTGATGGTATTC 1551
 Db 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValTyArgLysAsnValMetValPhe 480
 QY 1552 ACCGACTGGATTATCGACAATGAGCGGACGCGC 1587
 Db 481 ThrAspTrpIleTyArgGlnMetArgAlaAspGly 492

RESULT 9

AA92050

ID AA92050 standard; protein; 492 AA.

XX AC AA92050;

XX 01-AUG-2000 (first entry)

XX DE HrPCa6/7 polypeptide from androgen-inducible gene clone.

XX XX Androgen inducible; testosterone; prostate cancer; cytostatic; TMPRSS2;

XX KW diagnosis.

XX OS Homo sapiens.

XX PN W0200018961-A2.

XX XX 06-APR-2000.

XX PF 30-SRP-1999; 99WO-US022535.

XX PR 30-SRP-1998; 98US-00163759.

XX PR 30-SRP-1998; 98US-00164159.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Macbeth KJ, Shyjan AW;

XX DR WPI; 2000-293182/25.

XX DR N-PSDB; AAA08803.

XX XX Novel methods for identifying compounds for treating prostate cancer

XX PT comprising measuring the level of expression or activity of 1 or more of

XX PT 11 Genes or their products.

XX PS Claim 2; Fig 3; 108pp; English.

XX CC This protein is encoded by a gene which is androgen (e.g. testosterone)

XX CC inducible in androgen-dependent prostate cancer cells (e.g. LNCaP cells)

XX CC and constitutively expressed in androgen-independent prostate cancer

XX CC cells (e.g. LNCaP cells). Agents which decrease the expression or

XX CC activity of these clones may slow or arrest the growth of prostate cancer

XX CC cells or may kill them. HrPCa6/7 can be obtained from the sequence of the

XX CC known gene for TMPRSS2. A compound useful for treating prostate cancer

XX CC can be identified in a novel method comprising measuring the expression

XX CC level, or activity, of HrPCa2, 3, 6/7, 8, 9, 10, 13, 14, 15, 19, or

XX CC peripheral-type benzodiazepine receptor (PBR) in a cell, in the presence

XX CC and absence of a test compound. The sequences may also be used in

XX CC diagnosis of prostate cancer and to determine efficacy of treatment for

XX CC prostate cancer

XX SQ Sequence 492 AA;

Alignment Scores:

Pre. No.: 3,48e-221 Length: 492
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 85.22% Indels: 0

DB: 3 Gaps: 0
 US-09-323-597C-1 (1-1738) x AA92050 (1-492)
 QY 112 ATGGCTTTCAACTCAGGGGTCAACACAGCTATTGGACCTTACTATGAAAAACCATGGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyTrpGluAsnHsGlyTy 20
 QY 172 CAACCGGAAAAACCCCTATCCCGCAGCCACTGTGGTCCCCACTGTCTTACGAGGTGCAT 231
 Db 21 GlnProGluAsnProTyProAlaGlnProThrValValProThrValTyGluValHs 40
 QY 232 CCGGCTCAGTACTACCCGTCCTCCCGTCCCGTCCAGTACGCCCGAGGGTCTCTGACGAGGT 291
 Db 41 ProAlaGlnTyTrpProSerProValProGlnTyAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCCGTCTGTGCACGAGCCCAATCCCATCCGACAGTGTGCACCTCAAAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCCCTGGGAGCTTCTCTGTGGAGCTGGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 GCGGCTGGCTACTTGGAAAGTTCAGGCGCAGCAAGTGTCCCACTCTGGGATAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCTCAGGTACCTGCATCAACCCCTCTAACTGCTGTGATGGGTGTGCACCTGCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGCGGAGGACAGAAATCGGTGTGTGGCTTCTAGGACCAACACTTCTCTGAGTGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyGlySerAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGAGGAAGTCTGGCACCCTGTGTGTCAGACCACTGGAACCAAGACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTy 180
 QY 652 GCGGCGGCGCTGTGCAGGACATGGGCTATAAGAATAATTTTCTTAGCCCAAGAAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyTrpLysAsnAsnPheTyTrpSerGlnGlyIle 200
 QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAAAGTCCCGGCAATTCGAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACACAGTATGCTGTCTTCAAAAGCAGTGTCTTTTACGC 831
 Db 221 IleTyLysLysLeuTyHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTGGGGGTCACTTGAACCTCAAGCCCGCAGGAGGATTTCTGGCGGCGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGGCTCTCCCGGGGCGCTGGCCCTGGCAGGTCAGCTGCACCTCCAGAACCTCCACGTG 951
 Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TCGGAGGCTCCATCATACCCCGGAGTGTGATGAGTGTGACAGCCCGCCCTCTGGNAAA 1011
 Db 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTAACAATCCATGGCATTGGACGGCATTGGGGGATTCTTGAGACAATCTTTTCATG 1071
 Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATATGACTCCCAAG 1131
 Db 321 PheTyGlyAlaGlyTyTrpGlnValGluLysValIleSerHisProAsnTyTrpAspSerLys 340
 QY 1132 ACCAAGAACAATGACATTCGGCTGTGATGAGCTGCAGAGCCCTCTGACTTTTCAACGACCTA 1191


```
QY 892 AGCGCGCTCCCGGGGCTGGCCCTCGAGTGCAGCTGCACGTCCAGAACTCCACGTG 951
PR |||||
DB 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGCGAGGCTCCATCATCCACCCCGAGTGCATCGTGCAGCGCCGCCACTGCGTGGAA 1011
DB |||||
DB 281 CysGlySerIleIleTrpGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGCATTCGACGGCATTTGCGGGATTTTGAGACAATCTTTCATG 1071
DB |||||
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGACCGCATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATATGATCTCAAG 1131
DB |||||
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTCGCTGATGAAGCTCAGAGAGCTCTGACTTTTCAACGACCTA 1191
DB |||||
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACCATGTGTCTGCCCAACCCAGGATGATGCTGAGCCAGAACAGCTCTGCTGG 1251
DB |||||
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAGTGTGTGAACGTGCC 1311
DB |||||
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGCTCTCTATGACACAGAGATGCAACAGCAGATATCTCTATGACAACTGTATC 1371
DB |||||
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCAACCATGATCTGTGCGGCTCCTGCAGCGGAAAGTGCATCTCCAGGCTGAC 1431
DB |||||
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTGTGTCACCTTCGAGACAAATATCTGTGGTGTATGAGGGGATACAAGC 1491
DB |||||
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGCTGTGCGCAAGCTTACAGACCGAGCTGTACGGGAATGTGATGTTTC 1551
DB |||||
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACATGGATTTATCGACAAATGAGGGCAGACGGC 1587
DB |||||
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
RESULT 11
AAU69960
XX AC AAU69960 standard; protein; 492 AA.
XX AC AAU69960;
XX 30-JAN-2002 (first entry)
DE Human prostate cDNA encoded protein #86.
XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX Homo sapiens.
XX WO200173032-A2.
XX 04-OCT-2001.
XX 27-MAR-2001; 2001WO-US009919.
XX 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
```

```
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SK, Wang A, Skeiky YAW, Hepler WT, Henderson RA,
XX WPI; 2001-639232/73.
XX N-PSDB; AAS64178, AAS64179, AAS64180.
PR New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer.
XX Claim 2; Page 573-574; 579pp; English.
XX The invention relates to isolated prostate-specific polynucleotides,
CC polypeptides, fusion proteins of the polypeptides, antibodies raised
CC against the polypeptides (or antigenic epitopes derived from them) and
CC antigen-presenting cells expressing the polypeptides. The antibodies are
CC useful for detecting the presence of cancer, especially prostate cancer.
CC The polypeptides, polynucleotides and the antigen-presenting cells are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein, and for inhibiting the development of cancer especially prostate
CC cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC useful for stimulating an immune response, and for treating cancer. The
CC oligonucleotide is useful for detecting cancer. The present sequence is a
CC prostate specific polypeptide of the invention
XX Sequence 492 AA;
SQ
Alignment Scores:
Pred. No.: 3,48e-221 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.82% Indels: 0
DB: Gaps: 0
US-09-323-597C-1 (1-1738) x AAU69960 (1-492)
QY 112 ATGGCTTTGAAGTCAAGGCTCACCACGCTATTGGACCTTACTATGAAACCATGGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrIleGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCAGCCCACTGTGGTCCCCCACTGTCTACGAGGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCAACCCCGCTGTCTGTCAGCGAGCCCAATCCCATCCGGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCAGCTGTGCATCACCTTGACCTGGGAGACCTTCTCTGGGAGCTGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGGCTGGGCTACTCTCGGAAGTTTCATGGCAGCAAGTGTCTCAACTCTGGGATAGAGTC 471
DB 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCTCAGGTACCTGTGCATCAACCCCTCTAATGGTGTGATGGCGGTGTGCACACTGCC 531
DB |||||
```

Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGAGGAGCAGAGATCGGTGTCTTGGCTCTACGACCAAACTTCATCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAGAGTCTGGCAGCCTGTGTGCGCAGAGCTGGAGCAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GGGCGGGCGGCTGCGAGGACATGGCTATAAGATAATTTTACTCTAGCAAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGGATCACCAAGCTTGTATGAACACAGAGTCCGGGCAATGTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAGTGTACACAGATGATGCTTCTTCAAAAGCAGTGTGTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGATATAGCTGGGGTCAACTTGAATCAAGCCGACAGACAGATTTGTGGCGGCGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCGGGGCGCTGCGCGAGTGCAGCTGCAGTCCAGACGAGTCCAGAGTCCAGT 951
Db 261 SerAlaLeuProGlyValAspTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCGCGCCACTGCGTGGAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTACATCATGATGCGATGAGCGCATTTGGGGGATTTTGAGCAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCAAAATTTATGATCTCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGACATGATGCTGCGTATGAGCTGCGAGAGCTCTGAGCTTCAAGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCCAGTGTGTCTGCCAACCCAGGCTATGCTGCGAGCCAGACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCGCCAGAGAGAAAGGAGACCTCAGAGTGTCTGAACGCTGCC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTTGACACACAGAGATGCAACAGCATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACAGCCATGATCTGCGCGCTTCTGCGAGGGAACGTCGATTTCTGCCAGGCTCAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTGCTGCTACCTCGAAGAACATATCTGCTGCTGATGAGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGCTGTGCGAAAGCTTACAGACAGGAGTGTACGGGATGTGATGCTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGAGCTGATTTATCGAAATATGAGGCGCAGACGCT 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 12

ABU71851

ID ABU71851 standard; protein; 492 AA.

XX AC ABU71851;

XX DT 10-JUN-2003 (first entry)

XX DE Prostate cancer associated protein #67.

XX KW Prostate cancer; vaccine; Gene therapy; cytostatic; fusion protein;

XX KW immunogen; cancer; prostate specific antigen; PSA;

XX KW Prostatic acid phosphatase; PAP; prostate specific membrane antigen;

XX KW PSMA.

XX OS Homo sapiens.

XX PN US2002192763-A1.

XX PD 19-DEC-2002.

XX PF 29-JUN-2001; 2001US-00895793.

XX PR 04-OCT-1999; 99US-0157455P.

XX PR 04-OCT-2000; 2000US-00679272.

XX PR 28-MAR-2001; 2001US-00822827.

XX PA (XUJ)/ XU J.

PA (DILL)/ DILLON D C.

PA (MITC)/ MITCHAM J L.

PA (HARL)/ HARLOCKER S L.

PA (JIAN)/ JIANG Y.

PA (KALO)/ KALOS M D.

PA (FANG)/ FANGER G R.

PA (RETT)/ RETTER M W.

PA (STOL)/ STOLK J A.

PA (DAYC)/ DAY C H.

PA (VEDV)/ VEDVICK T S.

PA (CART)/ CARTER D.

PA (LISX)/ LI S X.

PA (WANG)/ WANG A.

PA (SKBI)/ SKBI Y A W.

PA (HEPL)/ HEPLER W T.

PA (HEND)/ HENDERSON R A.

PA (HURA)/ HURAL J.

PA (MCNE)/ MCNEILL P D.

PA (HOUT)/ HOUGHTON R L.

PA (DBAS)/ Y DE BASSOLS C V.

PA (FOYT)/ FOY T M.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

XX Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;

XX Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;

XX McNeill PD, Houghton RL, Y De BassolsCV, Foy TM;

XX NFI; 2001-245062/25.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?docID=US20020192763

SQ Sequence 492 AA;

Alignment Scores:

Pred. No.: 3 48e-221 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 4 Gaps: 0

US-09-323-597C-1 (1-1738) x ABU71851 (1-492)

QY 112 ATGGCTTGAACATCAGGCTCAGCAGCTATTGGACCTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGCCACTGTGTCCCTCCACTCTCTACGAGTGCAT 231
Db 21 GluProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGCTCCCGGTGCCCGAGTACGCCCGGGTCTCTACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCACCCCGCTCTCGACGAGCCCAATCCCATCCCGGACAGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAACACTGTGCATCACCTTGACCTCGGACCTTCTCTGGAGCTGGCGTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCTGGCTACTCGAAGTTCATCGGACGACAGTGTCCCACTTGGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACTGTCAACCCCTCAACTGTGTGTATCGGTGTGACGCTGACACTGCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGCAGAGATCGGTGTGCTCTTACGGACCAAACTTCATCTCTCGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAGTCTCGCACCTGTGTGCCAAGCACTGGAACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTCGCAGGACATGGCTATAGAAATAATTTTACTCTAGCCCAAGAAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATCAGACCGGATCCACCGTTTATGAACATGAACACAGTCCCGCATGTCTGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTACCAAGTGTGCTGTCTTCAAAAGCAGTGTCTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCACTTGAACCTCAAGCGCCAGACGAGATGTGGCGCGGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGCTGGCGCTGCGAGTGTGACGCTGACGCTCCAGAACTGCCACGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTTCATCATCAACCCCGAGTGTGATCGTACAGCCGCGCCCACTCGCTGGAAAAA 1011

Db 281 CysGlyGlySerIleIleThrProGluTyrPheIleValThrAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGCGATTTGGACGCGCATTTGGGGGATTTGAGACAACTTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGGCGGATACCAAGTACGAAAGGATTTCTCATCCCAATATGACTCTCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTTGCGCTGTAGCTGCGAGCTGCGAGACCTCTGACTTTTCAACGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACAGTGTCTGCGCAACCGCATGATGCTGCAGCCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAACTGCTGAACGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTATTGAGACACAGATGCGACAGCAGATATGTCTATCACACCTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACAGCCATCATCTGTGCGCGCTTCTCGAGGGGAACTCGATTCTTGCAGGCTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 ATGAGAGGCGCTCTGCTCACTTCGAGAACAAATATCTGCTGGCTGTAGAGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TCGGGTCTGCTGTCGCCAAAGCTTACAGACGAGGAGTGTACGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGGATTTATCGAAATGAGGCGACGCGC 1587
Db 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
RESULT 13
ID ABB95420
ID ABB95420 standard; protein; 492 AA.
XX AC ABB95420;
XX 19-JUL-2002 (first entry)
XX DE Human P1000C SEQ ID NO 932.
XX KW Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;
XX KW gene therapy.
XX OS Homo sapiens.
XX XX US2002022248-A1.
XX PD 21-FEB-2002.
XX PP 12-JAN-2001; 2001US-00759143.
XX PR 25-FEB-1997; 97US-00806099.
XX PR 01-AUG-1997; 97US-00904804.
XX PR 14-FEB-1998; 98US-00020956.
XX PR 25-FEB-1998; 98US-00030607.
XX PR 14-JUL-1998; 98US-00115453.
XX PR 23-SEP-1998; 98US-00159812.
XX PR 15-JAN-1999; 99US-00232149.
XX PR 09-APR-1999; 99US-00288946.
XX PR 13-JUL-1999; 99US-00352616.

PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 10-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
XX (XUJJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIAN/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISX/) LI S X.
PA (WANG/) WANG A.
PA (SKEI/) SKEIKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2002-255649/30.
XX
PT New prostate-specific polynucleotides for diagnosing and treating
PT diseases, in particular prostate cancer, and as markers for the
PT progression of cancer.
XX
PS Claim 2; SEQ ID NO 932; 87bp; English.
XX
CC The present invention provides prostate-specific coding sequences and
CC their encoded proteins. These can be used in the diagnosis and treatment
CC of cancers, particularly prostate cancer. The present sequence is a
CC protein described in the invention
XX
SQ Sequence 492 AA;
Alignment Scores:
Pred. No.: 3 48e-221 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 5 Gaps: 0
US-09-323-597C-1 (1-1738) x ABB95420 (1-492)
QY 112 ATGGCTTTGAACCTAGGCTCACCAGCTATTGGACCTTACTATGAAACCATGGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCAGACGACCTGTTGGTCCCACTGTCTACGAGTGCCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGTCCCGGTCGCCAGTACGCCCGAGGTCCTGACGAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60

QY 292 TCCAAACCCCGTCTGTGACGACGCCCAAAATCCCATCCGGGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACCTTGCACCTGGGACCTTCTCTCGTGGGAGCTGCGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCCGCTGGCTACTCTGGAAGTTTCATGGGACAGCAAGTGTCTCCAACTCTGGATAGAGTGC 471
DB 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGGTACCTGCATCAACCCCTCTAACTGTGTGATGGCGTGTGCACACTGCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGAAATCGGTGTGCTCTACGGACCAAACTTCACTCTTCAGCTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAAGTCTGGCCACCTGTGTGTCACAGACGACTGGACGAGAACTAC 651
DB 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTGCGAGGACATGGGCTATAAGATAATTTTACTCTAGCCAAAGGAATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
QY 712 GTGGATCAGACGCGATCCACGACTTTATGAACCTGAACACAGTCCGCGCAATGTGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACACAGTGTGCTGTCTTCAAAAGCAGGTGTTCTTTACGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
QY 832 TGTATAGCTGCGGGGTCAACTTGAACCTCAAGCCGCGCAGACGAGTGTGGCGGCGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
QY 892 AGCGCGTCCCGGGGCGCTGCGCTGCGAGCTGACCTGACCTCCAGACGCTCCAGCTG 951
DB 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCCGCCACTCCGTGGGAAAAA 1011
DB 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACCAATCCATGGCATTTGGACGGCATTTGCGGGGATTTTGAGACAATCTTTCATG 1071
DB 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATTTACTCCCAAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACCAATGACATTTGCGCTGATGAAGCTCAGAGGCTCTGACTTTCAACAGCCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCAAGTGTCTGCGCCCAACCCAGGCTCATGCTGACGACGAGCTCTGCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
QY 1252 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGGAAGACCTCAGAAGTGTCTGAACGCTGCC 1311
DB 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGCTTCTCATTTAGACACAGATGCCAACAGACAGATATCTCTATGACACCTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCGCATGATCTGTGCGCGCTTCTGACGGGAAAGCTGCTGATTTCTTCCAGGGGTGAC 1431

Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACGAGTGTCTGCCAACCCAGGACATGATGCTGCAGGCAGAACACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCGGGTGGGGCCACCGAG 1311
Db 381 IISerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCATGATCTGCGGCTCTCTGCGAGGAGACGTCGATCTTGCAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGAGGCGCTCTGCTCACTTCGAAGAACAAATATCTGCTGCTGATAGGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGCTGTCGCAAGCTTACAGACAGAGTGTACGGGATGTGATGCTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGATTATTCGACAAATGAGGCGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaaspGly 492
RESULT 15
ADBI4382
ID ADBI4382 standard; protein; 492 AA.
XX AC ADBI4382;
XX DT 18-DEC-2003 (first entry)
XX DE Human prostate specific protein P1000C.
XX KW Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
XX KW cell therapy; vaccine; T-cell epitope;
XX KW class I major histocompatibility complex allele; MHC; prostate cancer;
XX KW tumour; antigen presenting cell.
XX OS Homo sapiens.
XX PN US2003185830-A1.
XX PD 02-OCT-2003.
XX PF 12-NOV-2002; 2002US-00294025.
XX PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.

PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
PR 12-JAN-2001; 2001US-00759143.
PR 09-FEB-2001; 2001US-00780669.
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
PR 09-MAY-2002; 2002US-00144678.
XX (CORI-) CORIXA CORP.
XX PA Xu J, Stolk JA, Kalos MD;
XX PI WPI; 2003-756193/71.
XX DR N-PSDB; ADBI4379, ADBI4380, ADBI4381.
XX PT New isolated polypeptide for use in a vaccine for stimulating an immune
PT response, or for treating or diagnosis cancer, preferably prostate
PT cancer.
XX Example 2; Page; 101pp; English.
XX SS The invention relates to an isolated polypeptide comprising no more than
CC 11-542 amino acids of ADBI3563 comprising a sequence ADBI4487. The
CC peptides comprise a fragment ADBI3563 of that contain naturally processed
CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC alleles. ADBI3563 is a polypeptide encoded by a human prostate specific
CC cDNA one of 648 disclosed as new. Also included are nucleic acids
CC encoding the proteins and peptides, expression vectors, a host cell
CC transformed with the vector, an isolated antibody (or antigen binding
CC fragment) that specifically binds to the protein or peptide, detecting
CC the presence of a cancer in a patient (comprising contacting a patient
CC sample with a binding agent that binds to the peptides or a polypeptide
CC appearing as ADBI3558, detecting the amount of polypeptide that binds to
CC the agent and comparing the amount of polypeptide to a predetermined cut-
CC off value to determine the presence of cancer), a fusion protein
CC comprising the peptides or proteins, stimulating or expanding T cells
CC specific for a tumour protein comprising contacting T cells with the
CC peptides or the isolated T cell population, treating prostate cancer in a
CC patient comprising administering a composition comprising the peptides,
CC nucleic acids, antibodies or compounds, determining the presence of a
CC cancer in a patient and treating prostate cancer in a patient comprising
CC incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
CC from a patient with the peptides or antigen presenting cells that express
CC (the peptides so that the T cells proliferate, and administering the
CC proliferated T cells to the patient. The peptides (or an oligonucleotide
CC that hybridises to nucleic acid encoding them), is used to detect the
CC presence of cancer in a patient. The peptides, nucleic acids encoding, or
CC antigen-presenting cells expressing the nucleic acid, are used to
CC stimulate or expand T cells specific for a tumour protein. The peptides,
CC nucleic acids, antibodies, fusion proteins, T cell populations or antigen
CC presenting cells are used to stimulate an immune response or treat
CC prostate cancer in a patient. The present sequence is a prostate specific
CC protein of the invention. Note: Except where otherwise indicated, the
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=20030185830.
XX SQ Sequence 492 AA;

Alignment Scores:			
Pred. No.:	3,48e-221	Length:	492
Score:	2709.00	Matches:	491
Percent Similarity:	99.80%	Conservative:	0
Best Local Similarity:	99.80%	Mismatches:	1
Query Match:	85.22%	Indels:	0
DB:	7	Gaps:	0

US-09-323-597C-1 (1-1738) x ADBI4382 (1-492)

QY 112 ATGGCTTGAACCTCAGGCTCACCACGACTTGGACCTTACTATGAAACCATGGATAC 171
|||||

Db	1	MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
QY	172	CAACCGGAAAAACCCCTACTCCCGCACACCCCACTGTGTGTCCCCACTGCTTACGAGGTGCAT	231
Db	21	GlnProGluAsnProTyrProAlaIleProThrValValProThrValTyrGluValHis	40
QY	232	CGCGCTCAGTACTACCGTCCCGCTGCCCGCCAGTACGCGCCCGAGGTCCCTGACGCGAGCT	291
Db	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
QY	292	TCCAACCCCGCTCGCTCGACGACGCCCAAAATCCCCATCCCGGACAGTGTGCACCTCAAAG	351
Db	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
QY	352	ACTHAGAAAACGACTGTGCATCACTTGACCTGGGACCTTCCCTGGGAGTGCAGCTGCGCTG	411
Db	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
QY	412	CGCGCTGGCCTACTCTCGAAGTTCATGGGACAGCAAGTCTCCAACTCTGGGATAGAGTGC	471
Db	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
QY	472	GACTCTCAGGTACTCTGCATCAACCCCTCTAACTGTGTGTGATGCGTGCACACTGCCCC	531
Db	121	AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro	140
QY	532	GGCGGGAGACGAGAAATCGGTGTGTGCGCTCTACGACCAAACTTCATCTCTCAAGTG	591
Db	141	GlyGlyGluAspGluAsnAsCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal	160
QY	592	TACTCATCTCAGAGGAAGTCCTGCACCCCTGTGTGCCAAGACGACTCGAACGAGAACTAC	651
Db	161	TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr	180
QY	652	GGCGGGCGGCTTCAGGGACATGGGTATAGAAATATTTTACTCTACGCCAGGAATA	711
Db	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200
QY	712	GTGGATCACAGCGGATCCACGAGCTTTATGAAATCTGAACCAAGTCGCGGCAATGTCGAT	771
Db	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
QY	772	ATCTATAAAAACCTGTACCAGATGATGCTGTCTTCAAAGCAGTGGTTTCTTTACGC	831
Db	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
QY	832	TGTATACCTCTCGGGGTCAACTTCAACTCAAGCGCCAGACGAGATGTGGCGCGCGAG	891
Db	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu	260
QY	892	AGCGCGTCCCGGGGCGCTGGCCCTGGCAGGTCAAGCTGCAGTCCAGAACGTCACAGTG	951
Db	261	SerAlaLeuProGlyAlaIleTyrProTyrProGlnValSerLeuHisValGlnAsnValHisVal	280
QY	952	TGGCGAGGCTCCATCATCACCCCGAGTGATCTGTGACCGCCGACCTCGGTGGAAAAA	1011
Db	281	CysGlyGlySerIleIleThrProGluThrIleValThrAlaAlaHisCysValGluLys	300
QY	1012	CCTCTTAACCAATCCATGGCATTTGACGGCAATTTTCGGGGATTTTGTAGACAATCTTTCATG	1071
Db	301	ProLeuAsnAsnProTyrHisSerTyrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
QY	1072	TTCTATGGAGCCGATACCAAGTAGAAAGTGAATTTCTCATCCCAAAATTGACTCCAAG	1131
Db	321	PheTyrGlyAlaGlyTyrGlnValIleGluLysValIleSerHisProAsnTyrAspSerLys	340
QY	1132	ACCAAGAACCAATGATTCGCTCATCAAGCTGCAGAAAGCTCTGCACTTTCACACGACCTA	1191
Db	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
QY	1192	GTGAAAACAGTGTGTGTGCCCAACCCAGGCGATGATGTGTGAGCCAGAACAGCTCTGTGTG	1251
Db	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr	380

Search completed: July 6, 2004, 15:03:08
Job time : 108 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: July 6, 2004, 14:58:46 ; Search time 31.5 Seconds
(without alignments)
10614.659 Million cell updates/sec

Title: US-09-323-597C-1

Perfect score: 3179

Sequence: 1 ggcggaggcgagcgagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ r2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09323597/runat_06072004_111728_29674/app_query.fasta_1.1927
-DB=pir_78 -QWMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09323597 @CGN 1.1.46 @runat_06072004_111728_29674 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

pir_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	660	20.8	1035	A43090	enteropeptidase (E
2	647	20.4	1019	A56318	enteropeptidase (E
3	634	19.9	1034	A53663	enteropeptidase (E
4	579	18.2	1113	J03035	low-density lipopr
5	569.5	17.9	638	K0HUP	plasma kallikrein
6	567	17.8	417	S00845	hepsin (EC 3.4.21.
7	564.5	17.8	855	JC7731	membrane-bound arg
8	556	17.5	638	K0RTPL	plasma kallikrein
9	553.5	17.4	812	P1XMS	plasmin (EC 3.4.21
10	538	16.9	638	K0MSPL	plasmin (EC 3.4.21
11	533	16.8	625	K0PHU1	coagulation factor
12	531	16.7	416	K33777	hepsin (EC 3.4.21.
13	504.5	15.9	1524	T30337	polypeptidase - Afri
14	501	15.8	421	S11674	acrosin (EC 3.4.21

ALIGNMENTS

RESULT 1

A43090 enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine

N;Alternate names: enterokinase

C;Species: Bos primigenius taurus (cattle)

C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 28-Apr-2003

C;Accession: A43090; A48874; A61436

R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease con

A;Reference number: A43090; MUID:94329561; PMID:8052624

A;Accession: A43090

A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBU

A;Molecule type: mRNA

A;Residues: 1-1035 <KIT>

A;Cross-references: GB:U09859; NID:G746410; PIDN:AA040026.1; PID:G746411

A;Experimental source: small intestine

R;LaValle, E.R.; Rehentulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.

J. Biol. Chem. 268, 23311-23317, 1993

A;Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of

A;Reference number: A48874; MUID:94043122; PMID:8226855

A;Accession: A48874

A;Molecule type: mRNA

A;Residues: 801-1035 <LAV>

A;Cross-references: GB:U19663; NID:G416131; PIDN:AAA16035.1; PID:G416132

A;Note: parts of this sequence, including the amino end of the mature protein, were co

R;Light, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A;Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.

A;Reference number: A61436; MUID:92189715; PMID:1799406

A;Accession: A61436

A;Molecule type: protein

A;Residues: 801-807, 'Y', 809-827 <LIG>

C;Comment: The mechanism of association with the membrane of the intestinal brush borc

embrane attachment using a signal-anchor sequence.

C;Comment: Conversion from membrane-bound to soluble forms may involve further process

C;Complex: mature enteropeptidase is variously reported to contain two (heavy and light

lfiide linked

otated below) or with amino-terminal myristoylation of the heavy chain.

C:Genetics:
A:Gene: GDB:PRSS7
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light) ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv ducts.

C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding res
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <RCH>
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MAM homology <MAM>
F:526-631/Domain: C1r/C1s repeat homology <C1R>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRQ>
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Alignment Scores:

Pred. No.:	1.16e-38	Length:	1019
Score:	647.00	Matches:	138
Percent Similarity:	53.25%	Conservative:	67
Best Local Similarity:	35.84%	Mismatches:	154
Query Match:	20.35%	Indels:	26
DB:	1	Gaps:	10

US-09-323-597C-1 (1-1738) x A56318 (1-1019)

```

Qy 448 TGTCCAACTCTGGATAGAGTGGCGACTCTCAGTACCTGCATCAACCCCTCTACTGG 507
Db 643 CysLysAlaAspHisPheGlnCys---LysAsnGlyGluCysValProLeuValAsnLeu 661
Qy 508 TGTGATGCGCTGTACACTGCCCGCGGGAGGAGGAGAAATCGGTGTTCCTCCCTCTAC 567
Db 662 CysAspGlyHisLeuHisCysGluAspGlySerAspGluAlaAspCysValArgPhePhe 681
Qy 568 GGA-----CCAACTTCATCTCTCAGGTGTACTCATCTCAG 693
Db 682 AsnGlyThrThrAsnAsnAsnGlyLeuValArgPheArgileGln----- 696
Qy 604 AGGAAGTCTTGGCACCCTGTGTGCCAAGACGACTGGAAACGAACTACGGCGGGCGGCC 663
Db 697 ---SerileThrHisThrAlaCysAlaGluAsnTrpThrGlnIleSerAsnAspVal 715
Qy 664 TGCAGGACATGGGCTATAAATAATATTTTACTTACCAAGAAATAGTGGATGACAGC 723
Db 716 CysGlnLeuLeuGlyLeuGlySerGly---AsnSerSerLysProIlePheSerThrAsp 734
Qy 724 GGAATCCACAGCTTTATGAACATGACACAGAGTCCGGCAATGCGATCTATATAAAA 793
Db 735 GlyGly---ProPheValLysLeuAsnThrAlaPro-----AspGlyHisLeuLeu 750
Qy 784 CTGTACACAGTGTGCTGTCTTCAAAAGCAGTGTCTTTCTTAGCTGT-----ATA 837
Db 751 LeuThrProSerGlnGlnCysLeuGlnAspSerLeuLeuArgLeuGlnCysAsnHisLys 770
Qy 838 GCTTGGGGTCACTTGAATCAAGCCGC---CAGAGCAGGATTTGGGGCGGAGAGC 894
Db 771 SerCysGlyLysLysLeuAlaAlaGlnAspIleThrProLysIleValGlyGlySerAsn 790
Qy 895 GCGTCCCGCGGGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 954
Db 791 AlaLysGluGlyAlaTrpProTrpValValGlyLeuTyrtYrGlyGlyArgLeuLeuCys 810
Qy 955 GGAGGGTCCATCATACCCCGGAGTGGATCTGTGACAGCGCCCGCATCTGGTGGAAAACCT 1014

```

```

Db 811 GlyAlaSerLeuValSerSerAspTrpLeuValSerAlaAlaHisCysValTyrGlyArg 830
Qy 1015 CTTAACAATCCATGGCATTGGCGGATTTGGCGGATTTGGAGACAATTTCTTCATGTT 1074
Db 831 AsnLeuGluProSerLysTrpThrAlaIleLeuGlyLeuHisMetLysSerAsnLeuThr 850
Qy 1075 -----TATGGACCGGATACCAAGTAGAAAAGTGTCTCTCATCAATTTGACTCC 1128
Db 851 SerProGlnThrValProArgLeuIleAspGluIleValIleAsnProHisTyrAsnArg 870
Qy 1129 AAGACCAAGAACAATGACATTGGCGTGCATGAAGCTGCAGAAGCTCTGACTTTCAACGAC 1188
Db 871 ArgArgLysAspAsnAspIleAlaMetMetHisLeuGluPheLysValAsnTyrThrAsp 890
Qy 1189 CTAGTGAACCAAGTGTCTGCCCCAACCCAGGATGCTGCGAGCCAGACAGCTCTGC 1248
Db 891 TyrIleGlnProIleCysLeuProGluGluAsnGlnValPheProGlyArgAsnCys 910
Qy 1249 TGGATTTCGGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAGTGTCTGAAACGCT 1308
Db 911 SerIleAlaGlyTrpGlyThrValValTyrGlnGlyThrThrAlaAsnIleLeuGlnGlu 930
Qy 1309 GCCAAGTGCTCTCATTTGAGACACAGATGCAACAGCAGATATGTCTATGACACCTG 1368
Db 931 AlaAspValProLeuLeuSerAsnGluArgCysGlnGlnGlnMetProGluTyrAsn--- 949
Qy 1369 ATCACACAGGACATGATCTGTGCGGCTCTCTGACAGGAAACCTGCTCTGCCAGGGT 1428
Db 950 IleThrGluAsnMetIleCysAlaGlyTyrGluGluGlyGlyIleAspSerCysGlnGly 969
Qy 1429 GACAGTGGAGGCGCTCTGCTACTTCGAAAGAACAAATATCTGGTGGCTGTAGGGGATACA 1488
Db 970 AspSerGlyGlyProLeuMetCysGlnGluAsnAsnArgTrpPheLeuAlaGlyValThr 989
Qy 1489 AGCTGGGGTCTCGCTGCTGCAAGCTTACACAGCAGGAGTGTACCGGATGTGATGCTGA 1548
Db 990 SerPheGlyTyrLysCysAlaLeuProAsnArgProGlyValTyrAlaArgValSerArg 1009
Qy 1549 TTCACGAGCTGGATT 1563
Db 1010 PheThrGluTrpIle 1014

RESULT 3
A53663
enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N:Alternate names: enteropeptidase
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C:Accession: A53663
R:Watanabe, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa
J. Biol. Chem. 269, 19976-19982, 1994
A:Title: Structural characterization of porcine enteropeptidase.
A:Reference number: A53663; PMID:94327548; PMID:8051081
A:Accession: A53663
A:Molecule type: mRNA
A:Residues: 1-1014 <MAT>
A:Cross-references: GB:ID30799; NID:9505122; PID:BAA06459.1; PID:9505123
A:Comment: Parts of this sequence, including the amino ends of three chains isolated from
otated below) or with amino-terminal myristoylation of the heavy chain.
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAM homology <MAM>

```

Alignment Scores:		
Pred. No.:	1e-37	Length: 1034
Score:	634.00	Matches: 147
Percent Similarity:	49.55%	Conservative: 72
Best Local Similarity:	33.26%	Mismatches: 179
Query Match:	19.94%	Indels: 44
DB:	1	Gaps: 13

929 GlyTrpGlyLysValIleTyrGlnGlySerProAlaAspIleLeuGlnGluAlaAspVal 948

Alignment Scores:	
Pred. No.:	9.07e-34
Score:	579.00
Length:	1113
Matches:	138
Percent Similarity:	48.58%
Conservative:	68
Best Local Similarity:	32.55%
Mismatches:	162
Query Match:	18.21%
Indels:	56
DB:	2
Gaps:	15
US-09-323-597C-1 (1-1738) x J080315 (1-1113)	

US-09-323-597C-1 (1-1738) x KOHUP (1-638)

```
QY 274 AGGTCCTCAGCAGGCTTCCAAACCCGCGTGTGTCAGCAGCAGCCAAATCCCATCCGGG 333
Db 218 ArgValLeuThr-----ProAspAlaPheValCysArg-----228
QY 334 ACAGTGTGACCTCAAGAGCTAAGAAGACACTGTGCATCACTTACCTGGGACCTTC 393
Db 229 ThrIleCysThrTyrHisProAsn-----CysLeuPhePheThrPheTyrThr---244
QY 394 CTGTGGGAGCTGCGGTGCGCGTGTGCTTCTGGAAGTTTCATGGGCAGC-----444
Db 245 -----AsnValTyrLysIleGluSerGlnArgAsnVal 255
QY 445 -----AAGTGTCCAACTCTGGATAGAGTGGACTCC-----477
Db 256 CysLeuLeuLysThrSerGluSerGlyThrProSerSerSerThrProGlnGluAsnThr 275
QY 478 -----TCAGGT-----ACCTGCATCAACCC-----TCTAAC 504
Db 276 IleSerGlyTyrSerLeuLeuThrCysLysArgThrLeuProGluProCysHisSerLys 295
QY 505 TGGTGTGATGGGTGTACACTGCCCGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
Db 296 IleTyrProGlyValAspPhe-----GlyGlyGluGluLeuAsn-----308
QY 565 TACGGACCAAACTTCACTCTCAGGTGTACTCATCTCAGAGGAGAGAGAGAGAGAGAG 624
Db 309 -----ValThrPheValLysGlyVal-----AsnVal 317
QY 625 TCCAGAGACACTGGAAACAGAGAACTACGGCGGGCGGCGGTGTCAGGAGCATGGCTAT 684
Db 318 CysGlnGlu-----ThrCysThrLysMet-----IleArg 327
QY 685 AATAATTTTACTAGCCAGGAATAGTGGATGACAGCGGATCAGCGGATCCACAGC-----735
Db 328 CysGlnPhePheThrTyrSerLeuLeuProGluAspCysLysGluGluLysCysLysCys 347
QY 736 TTATGAACTGAACACACAGTCCCGGCAATGTCGATATCTATAAAAACTGTACACACAGT 795
Db 348 PheLeuArgLeuSerMetAspGlySerProThrArgIle-----AlaTyrGlyThr 364
QY 796 GATGCTGTCTTCAAAAGCACTGTTCTTAACCTGTATAGCTTGGGGGTCAAC-----852
Db 365 GlnGlySerSerGly-----TyrSerLeuArgLeuCysAsnThrGlyAspAsnSer 381
QY 853 ---TTGAACCTCAAGCCGACAGACAGGATTGGGGCGGCGAGAGCGGCTCCCGGGGCC 909
Db 382 ValCysThrThrLysThrSerThrArgIleValGlyGlyThrAsnSerSerTyrGlyGlu 401
QY 910 TGGCCTGGCAGCTGACCTGCAGCTCCAG-----AAGTCCAGCTGTGGGAGGC 960
Db 402 TrpProTyrGlnValSerLeuGlnValLysLeuThrAlaGlnArgHisLeuCysGlyGly 421
QY 961 TCCATCATCACCCCGAGTGGATCTGTGACAGCGCCGCTGCTGGA---AAACCTCTT 1017
Db 422 SerLeuIleGlyHisGlnTyrValLeuThrAlaHisCysPheAspGlyLeuProLeu 441
QY 1018 AACATCCATGCATGACAGCGCATTTGGCGGATTTTGAGCAATCTTTCATGTTCTAT 1077
Db 442 GlnAspValTyrArg-----IleTyrSerGlyIleLeuAsnLeuSerAspIleThrLys 459
QY 1078 GGAGCGGATAC---CAAGTAGAAAAGTATTTCTCATCCAAATATGACTCCAGACC 1134
Db 460 AspThrProPheSerGlnIleLysGluIleIleHisGlnAsnTyrLysValSerGlu 479
QY 1135 AAGAACATATGATTCCTGATGAGCTGCAAGAGCTCTGACTTTCACAGCACTAGT 1194
Db 480 GlyAsnHisAspIleAlaLeuIleLysLeuGlnAlaProLeuAsnTyrThrGluPheGln 499
QY 1195 AAACAGTGTGTCTCCCAACCCAGCATGATGCTGACAGCAGACACTCTGTGGATT 1254
Db 500 LysProIleCysLeuProSerLysGlyAspThrSerThrIleTyrThrAsnCysTrpVal 519
```

```
QY 1255 TCCGGTGGGGGCCACCCAGGAGAAAGGAGACCTCAGAAAGTGTCTGAACGCTGCCAAG 1314
Db 520 ThrGlyTyrGlyPheSerLysGluLysGlyGluIleGlnAsnIleLeuGlnLysValAsn 539
QY 1315 GTGCTTCTCATTTGAGACACAGATGCAACAGCAGATATGCTATGACAACTGATCACA 1374
Db 540 IleProLeuValThrAsnGlnGluCysGlnLysArgTyr---GlnAspTyrLysIleThr 558
QY 1375 CCAGCCATATCTGTGCGGCTTCTCTGACGGGAAACGTCGATTCTTCCAGGGTGACAGT 1434
Db 559 GlnArgMetValCysAlaGlyTyrLysGluGlyGlyLysAspAlaCysLysGlyAspSer 578
QY 1435 GGAGGGCTCTGCTCACTTCGAGAACAAATCTGGTGGCTGATAGGGGATACAGCTGG 1494
Db 579 GlyGlyProLeuValCysLysHisAsnGlyMetTyrArgLeuValGlyIleThrSerTyr 598
QY 1495 GGTTCCTGCTGTGCAAAAGCTTACAGACACAGGAGTGTACGGGATGTGATGTTTACG 1554
Db 599 GlyGluGlyCysAlaArgGluGlnProGlyValTyrThrLysValAlaGluTyrMet 618
QY 1555 GACTGGATTTATCGACAAATGAGG---GCAGACGCG 1587
Db 619 AspTyrIleLeuGluLysThrGlnSerSerAspGly 630
RESULT 6
S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C:Accession: S00845
R:Levy, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane
A:Reference number: S00845; MUID:88209431; PMID:2835076
A:Accession: S00845
A:Molecule type: mRNA
A:Residues: 1-417 <LEV>
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
C:Genetics:
A:Gene: GDB:HBN; TMPRSS1; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:23-45/Domain: transmembrane #status predicted <TMN>
F:163-400/Domain: trypsin homology <TRY>
F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted
Alignment Scores:
Pred. NO.: 6.23e-33 Length: 417
Score: 567.00 Matches: 143
Percent Similarity: 47.59% Conservative: 64
Best Local Similarity: 32.87% Mismatches: 167
Query Match: 17.84% Indels: 61
DB: 1 Gaps: 11
US-09-323-597C-1 (1-1738) x S00845 (1-417)
QY 340 TGACCTCAAGACTAAGAAAGCACTGTGCATCACCTTGACCTGGGGACCTTCTCTGTG 399
Db 12 CysCysSerArgProLysValAla-----AlaLeuThrAlaGlyThrLeuLeuLeu 28
QY 400 GGAAGCTCGCTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 459
Db 29 LeuThrAlaIleGlyAla-Ala-----SerTyrAlaIleValAlaValLeuLeu 44
QY 460 GGGATAGAGTGGCACTCTCTCAGTACCTGCATCAACCCCTCTAACTGGTGTGATGGGTG 519
Db 44 uArgSerAspGlnGluProLeuTyrProValGln-----55
QY 520 TCACACTGCCCGCGGGGAGGACGAGAAATCGGTGTGTTCGCTCTTACGGACCAAACTTC 579
```



```
QY 579 CATCTCTCAGGTGTACTCATCTCAGAGGAAGTCTGGCACCCCTGTGTGC----- 627
Db 500 rProCysGlnGlyTrpAlaGlnGluProHisArgHisSerIlePheThrProGlnTh 520
QY 628 -----CAAGACGACTGACAGCACTAGCGGGCGCGCCCTCGACGACATGGCC-- 678
Db 520 rAsnProArgAlaAspLeuGluLysAsnTrpCysArgAsnProAspGlyAspValAsnG 540
QY 679 -----TATAGATAATAATTTTACTCTAGCCAAAGTAATAGTGGATGACAGCGGATC 728
Db 540 yProTrpCysTrpThrThrAsn----- 547
QY 729 CACCACTTATGAACCTGACACAGAGTCCGCGCAATCTCGATATCTATAAAACTGTA 788
Db 548 -----ProArgLysLeuTy----- 552
QY 789 CCACAGTGATGCTGCTCTTCAAAAGCAGTGGTCTTACCTGCTATAGCTCGGGGGT 848
Db 552 r-----AspTyrCysAspIleProLeuCysAlaSerAlaSerSerPheGluCysGly 570
QY 849 CAACTTGAACTCAAGCCGCCAG-----AGCAGATGTGGCGCGGAGACGGCTCCC 902
Db 570 sProGlnValGluProLysLysCysProGlyArgValValGlyGlyCysValAlaAsnPr 590
QY 903 GGGGCGCTGGCCCTCGCAGGTCCAGCTGCAGCTCCAG-----AACGTCCACGTGTGGG 956
Db 590 ohISerTrpProTrpGlnIleSerLeuArgThrArgPheThrGlyGlnHisPheCysG 610
QY 957 AGGCTCCATCATACCCCGAGTGTGATCGTACAGCCGCCACCTCGCGGAAAAACCTCT 1016
Db 610 yGlyThrLeuIleAlaProGluTrpValLeuThrAlaAlaHisCysLeuGluLysSer-- 629
QY 1017 TAACAATCCATGCGCATTCGCGGCGATTTGGCGGATTTTGAGACATCTTTTCATGTTCTA 1076
Db 630 -SerArgProGluPheTrpLysValIleLeuGlyAlaHisGluGlu-----Ty 645
QY 1077 TGGAGCGGATACCAAGTAGAA-----AAAGTGATTTCTCATCAAAATTA 1121
Db 645 rIleArgGlyLeuAspValGlnGluIleSerValAlaLysLeuIleLeuGluProAsn-- 664
QY 1122 TGACTCCAGACCAAGACATGATTCGCTGATGAGCTGAGCTGAGAGCCCTTGACTTT 1181
Db 665 -----AsnArgAspIleAlaLeuLeuLysLeuSerArgProAlaThrIl 679
QY 1182 CAACGACCTAGTGAACAGGTGTCTGCCCAACCCAGCGATGATGTCGACCCAGACA 1241
Db 679 eThrAspLysValIleProAlaCysLeuProSerProAsnTrpMetValAlaAspArgTh 699
QY 1242 GCTCTGCTGGATTCCTGGGTGGGGGCCACCGAG-----GAGAAAGGGAAGACCTC 1292
Db 699 rIleCysTrpIleThrGlyTrpGlyGluThrGlnGlyThrPheGlyAlaGlyArgLeuLy 719
QY 1293 AGAAGTGTGAACGCTGCCAAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATA 1352
Db 719 sGlu-----AlaGlnLeuProValIleGluAsnLysValCysAsnArgValG 735
QY 1353 TGCTATGACAACTGATATCACACGACCATGATGTCGCGCTTCCTGCGGGGACGTT 1412
Db 735 uTyrLeuAsnAsnArgValLysSerThrGluLeuCysAlaGlyGlnLeuAlaGlyGlyVa 755
QY 1413 CGATTCTGCGAGGTGACATGAGCGGCTCTGCTGCTACTTCCGAGAACATATCTGCTG 1472
Db 755 lAspSerCysGlnIleAspSerGlyGlyProLeuValCysPheGluLysAspLysTrpIl 775
QY 1473 GCTGATAGGGGATCAAGCTCGGGTCTGCGTGTGCCAAAGCTTACAGACCAAGGAGTGA 1532
Db 775 eLeuGlnGlyValThrSerTrpGlyLeuGlyCysAlaArgProAsnLysProGlyValTy 795
QY 1533 CGGGAATGCTGATTCACGAGCTGATTTATGCAAAATGAGG 1578
Db 795 rValArgValSerArgPheValAspTrpIleGluArgGluMetArg 810
```

```
RESULT 10
KQMSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A36557
R:Seidman, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Roche
DNA Cell Biol. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compar
A:Reference number: A36557; MUID:91090844; PMID:2264928
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: GB:M58588; NID:G200358; PID:AAA63393.1; PID:G200359
A:Note: part of this sequence, including the amino ends of both the heavy and light c
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
are linked by one or more disulfide bonds.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
F:110-199/Domain: apple repeat <AP1>
F:200-289/Domain: apple repeat <AP2>
F:291-380/Domain: apple repeat <AP3>
F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F:31-621/Domain: trypsin homology <TRY>
F:127,215,308,396,494/Binding site: carbohydrate (asn) (covalent) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Alignment Scores:
Pred. No.: 7,78e-31 Length: 638
Score: 538.00 Matches: 115
Percent Similarity: 53.87% Conservative: 52
Best Local Similarity: 37.10% Mismatches: 117
Query Match: 16.92% Indels: 26
DB: 1 Gaps: 6

US-09-323-597C-1 (1-1738) x KQMSPL (1-638)
QY 664 TGACGAGCATGGGCTATAGATATATTTTACTTACCGCAAGATAGTGGATGACGC 723
Db 340 CysLysGluGluGlyCysLysCysSerLeuArgLeuSer-----ThrAspGlySer 356
QY 724 GGATCCACACAGCTTTATGAACTGAAACACACAAAGTCCGCGCAATCTCGATATCTATAAAAA 783
Db 357 ProThrArgIleThrTyrGlyMetGlnGlySerSerGlyTyrSerLeuArgLeuCysLys 376
QY 784 CTGTACACAGTGCCTGCTTCTTCAAAAGCAGTGGTTCCTTTACGCTGTATAGCCTGC 843
Db 377 LeuValAspSerProAspCysThrThrLysIle----- 387
QY 844 GGGGTCAACTGAACCAACGCGCAGAGCAGGATTTGGCGCGCGAGAGCGCGCTCCCG 903
Db 388 -----AsnAlaArgIleValGlyGlyThrAsnAlaSerLeu 399
QY 904 GGGGCTTCGCCCTCGGAGGTCACCTGCGTCCAGACAGTC-----CACGTGTGC 954
Db 400 GlyGluTrpProTrpGlnValSerLeuGlnValLysLeuValSerGlnThrHisLeuCys 419
QY 955 GAGAGCTCCATCATACCCCGAGTGGATCGTACAGCCGCCACCTCGGTGGAAAAACCT 1014
Db 420 GlyGlySerIleIleGlyArgGlnTrpValLeuThrAlaAlaHisCysPheAsp---Gly 438
QY 1015 CTTAAACAATCCATGGCATTTGGACGGATTTGGCGGGATTTTGAGACATCTTTTCATGTT 1074
Db 439 IleProTyrProAspValTrpArgIleTyrGlyIleLeuSerLeuSerGluIleThr 458
QY 1075 TATGGAGCC---GGATACCAAGTAGAAAAGTATTTCTCATCAATATGACTCCAAG 1131
Db 459 LysGluThrProSerSerArgIleLysGluLeuIleHisGlnGlyTrpLysValSer 478
```


QY 1132 ACCAAGACATGATCGCTGATGAGTCGCGAGAGCTCTGACTTTCACGACCTA 1191
Db 479 GlucylAsnTyrAspIleAlaLeuIleLysLeuInThrProLeuAsnTyrThrGluPhe 498
QY 1192 GTGAACACAGTGTCTGCTGCCAACCCAGGCATGATGTCGACGACCAACAGCTCTGCTGG 1251
Db 499 GlnLysProIleCysLeuProSerIleAlaAspThrAsnThrIleTyrThrAsnCysTyr 518
QY 1452 ATTTCGGTGGGGGGCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311
Db 519 ValThrGlyTyrGlyTyrThrLysGluGlnGlyGlnGlnGlnGlnGlnGlnGlnGln 538
QY 1312 AAGGTGCTTCATTTGACACAGAGATGACAGAGATGATGATGATGATGATGATGATGAT 1371
Db 539 ThrIleProLeuValProAsnGluCysGlnLysTyr---A-gAspTyrValIle 557
QY 1372 ACACAGCCATGATCTGCGGCTCTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
Db 558 AsnLysGlnMetIleCysAlaGlyTyrLysGluGlyGlyThrAspAlaCysLysGlyAsp 577
QY 1432 AGTCGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1491
Db 578 SerGlyGlyProLeuValCysLysHisSerGlyArgTrpGlnLeuValGlyIleThrSer 597
QY 1492 TGGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
Db 598 TrpGlyGluGlyCysGlyArgLysAspGlnProGlyValTyrThrLysValSerGluTyr 617
QY 1552 ACGACTGGATTATCGCAATAGGGCA 1581
Db 618 MetAspTrpIleLeuGluLysThrGlnSer 627

RESULT 11

KFHUI

Coagulation factor X1a (EC 3.4.21.27) precursor [validated] - human

A:Alternate names: antihemophilic factor C; plasma thromboplastin antecedent

C:Species: Homo sapiens (man)

C:Date: 13-Aug-1986 #sequence revision 26-May-1994 #text_change 08-Dec-2000

C:Accession: A27431; A00920; A37940

R:Asakai, R.; Davie, E.W.; Chung, D.W.

Biochemistry 26, 7221-7228, 1987

A:Title: Organization of the gene for human factor XI.

A:Reference number: A27431; MUID:88107663; PMID:2827746

A:Accession: A27431

A:Molecule type: DNA

A:Residues: 1-625 <ASA>

A:Cross-references: GB:M18295

A:Note: the sequence shown follows the authors' translation

R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.

Biochemistry 25, 2417-2424, 1986

A:Title: Amino acid sequence of human factor XI, a blood coagulation factor with four ty

A:Reference number: A00920; MUID:86243360; PMID:3636155

A:Accession: A00920

A:Molecule type: mRNA

A:Residues: 1-625 <FLU>

A:Cross-references: GB:M13142; MUID:g182832; PIDN:AAA52487.1; PID:g182833

R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.

Biochemistry 30, 2056-2060, 1991

A:Title: Location of the disulfide bonds in human coagulation factor XI: the presence of

A:Reference number: A37940; MUID:91152017; PMID:1998667

A:Accession: A37940

A:Molecule type: protein

A:Residues: 28-3335-49, 'X', 51-55, 'X', 57-63, 70-75, 'X', 77-79, 107-109, 'X', 111-112, 132-139,

280-282, 'X', 284, 285-297, 313-316, 'X', 318-319, 320-326, 'X', 328-330, 'X', 347-349, 373, 'X', 375

C:Comment: The proenzyme consists of two identical chains linked by one or more disulfid

he active site, and a heavy chain, which associates with high molecular weight (HMW) kin

C:Genetics:

A:Gene: GDB:E11

A:Cross-references: GDB:119891; OMIM:264900

A:Map position: 4q35-q435

A:Introns: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 52

C:Function:

A:Description: catalyzes the proteolytic activation of coagulation factor IX
A:Pathway: blood coagulation intrinsic pathway
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hy
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-387/Product: coagulation factor X1a heavy chain #status experimental <HCH>
F:19-108/Domain: apple repeat <AP1>
F:109-198/Domain: apple repeat <AP2>
F:199-288/Domain: apple repeat <AP3>
F:230-379/Domain: apple repeat <AP4>
F:388-625/Product: coagulation factor X1a light chain #status experimental <LCH>
F:388-618/Domain: trypsin homology <TRY>
F:20-103,514-581,571-595/Disulfide bonds: #status predicted
F:29/Disulfide bonds: interchain #status experimental
F:46-76,50-56,110-193,136-165,140-146,200-283,226-255,230-236,291-374,317-346,321-327
F:90,126,353,450/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:339/Disulfide bonds: interchain #status predicted
F:387-388/Cleavage site: Arg-Ile (coagulation factor X1a) #status experimental
F:431,480,575/Active site: His, Asp, Ser #status predicted
F:491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Alignment Scores:

Pred. No.: 1.78e-30 Length: 625
Score: 539.00 Matches: 122
Percent Similarity: 46.25% Conservative: 63
Best Local Similarity: 30.50% Mismatches: 129
Query Match: 16.77% Indels: 86
DB: 1 Gaps: 11

US-09-323-597C-1 (1-1738) x KFHUI (1-625)

QY 616 CACCTCTGTGTC-----CAAGACACTGGAGCAAGCACTACGGCGG 657
Db 233 HisProGlyCysLeuPhePheThrPheSerGlnGluTrpProLysGluSerGluArg 252
QY 658 GCGGCGCTGCGAGGACATGGGCTATAGATAATTTTACTCTAGCCAGCAATAGTCGAT 717
Db 253 AsnLeuCysLeuLysThrSerGluSerGlyLeuProSerThrArg-----IleLys 270
QY 718 GACAGCGGATCCACAGCTTTATGAACCTGACACAGTCGCCGCAATGTCGATATCTAT 777
Db 271 LysSerLysAlaLeuSerGlyPheSerLeuGlnSerCysArgHisSerIleProValPhe 290
QY 778 -----AARAACTGTACACACT----- 795
Db 291 CysHisSerSerPheTyrHisAspThrAspPheLeuGlyGluGluLeuAspIleValAla 310
QY 796 -----GATGCTGTTCTTCAAAAGCAGTGGTTTCTTTTACGCTTATA----- 837
Db 311 AlalysSerHisGluAlaCysGlnLysLeuCysThrAsnAlaValArgCysGlnPhePhe 330
QY 838 -----GCCTCGCGGGTCAACTTG 855
Db 331 ThrTyrThrProAlaGlnAlaSerCysAsnGluGlyLysGlyCysTyrLeuLysLeu 350
QY 856 AACTCAAGCGCG----- 867
Db 351 SerSerAsnGlySerProThrLysIleLeuHisGlyArgGlyGlyIleSerGlyTyrThr 370
QY 868 -----CAGACGAGATTGTCGGC 885
Db 371 LeuArgLeuCysLysMetAspAsnGluCysThrThrLysIleLysProArgIleValGly 390
QY 886 GGGGAGAGCGCGCTCCCGGGGCGCTGCGCTGCGAGTCCAGCTGCGACGTCACGACAGCTC 945
Db 391 GlyThrAlaSerValArgGlyGluTrpProTrpGlnValThrLeuHisThrThrSerPro 410
QY 946 -----CACGTGCGGAGGCTCATCATCACCCTCCGAGTGGATGCGACAGCGCGC 996
Db 411 ThrGlnArgHisLeuCysGlyGlySerIleIleGlyAsnGlnTrpIleLeuThrAlaAla 430
QY 997 CACTGGGTGGAACAACTCTTAACATCCATCCATGGACGCGCATTTGGGGGATTG 1056
Db 997 CACTGGGTGGAACAACTCTTAACATCCATGGACGCGCATTTGGGGGATTG 1056

Db 431 HisCysPhe---TyrGlyValGluSerProLysIleLeuArgValTyrSerGlyIleLeu 449
QY 1057 AGACAATCTTTTCATG-----TTCTATGAGCGCGATACCAAGTACAAAA 1101
Db 450 AsnGlnSerGluIleLysGluAspThrSerPhePheGly-----ValGlnGlu 465
QY 1102 GTGATTTCTATCCAAATATATGACTCCAGACCAAGAACAAATGACATTCGCTGATGAG 1161
Db 466 IleIleIleHisAspGlnTyrLysMetAlaGluSerGlyTyrAspIleAlaLeuLeuLys 485
QY 1162 CTGCAAGACCTCTGACTTCTCAACGACCTAGTGAACACAGTGTGTCTGCCAACCCAGGC 1221
Db 486 LeuGluThrThrValAsnTyrThrAspSerGlnArgProIleCysLeuProSerLysGly 505
QY 1222 ATGATGCTGCACCCAGAACAGCTCTGCTGATTTCCGGTGGGGGCCACCGAGAGAA 1281
Db 506 AspArgAsnValIleTyrThrAspCysTrpValThrGlyTyrGlyTyrArgLysLeuArg 525
QY 1282 GGGAGACCTCAGAAAGTGTGAACGCTGCCAAGGTGCTTCTCATGAGACACAGAGATGC 1341
Db 526 AspLysIleGlnAsnThrLeuGlnLysAlaLysIleProLeuValThrAsnGluGluCys 545
QY 1342 AACACAGATATGTTATGACAACTGATCAGACAGCCATGATCTGTGCGGCTTCCTG 1401
Db 546 GlnLysArgTyr---ArgGlyHisLysIleThrHisLysMetIleCysAlaGlyTyrArg 564
QY 1402 CAGGGGAAGCTGTGATCTTCCAGGGTGCAGCTGAGCGGCTCTGCTCACTTCGAGAAC 1461
Db 565 GluGlyGlyLysAspAlaCysLysGlyAspSerGlyGlyProLeuSerCysLysHisAsn 584
QY 1462 AATATCTGCTGCTGATAGGGGATACAACTGGGTTCTTGGCTGTGCCAAAGCTTACAGA 1521
Db 585 GluValTrpHisLeuValGlyIleThrSerTrpGlyGluGlyCysAlaGlnArgGluArg 604
QY 1522 CCAGAGCTACGGGAATGTGATGTATTCACGGACTCGATTCGACAAATGAGGCA 1581
Db 605 ProGlyValTyrThrAsnValValGluTyrValAspTrpIleLeuGluLysThrGlnAla 624

RESULT 12

hpsin (BC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; S32013
R:Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A:Title: Cloning and sequence analysis of rat hpsin, a cell surface serine proteinase.
A:Reference number: S33777; PMID:93305733; PMID:8318546
A:Accession: S33777
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <PAR>
A:Cross-references: ENBL:X70900; NID:g57928; PIDN:CAA50256.1; PID:g57929
C:Superfamily: hpsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status predicted <TM>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,290-358,321-337,348-380/Dissulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted

Alignment Scores:
Pred. No.: 2,418-30 Length: 416
Score: 531.00 Matches: 135
Percent Similarity: 44.64% Conservative: 69
Best Local Similarity: 29.54% Mismatches: 183
Query Match: 16.70% Indels: 70
DB: 1 Gaps: 11
US-09-323-597c-1 (1-1738) x S33777 (1-416)
QY 340 TCACCTCAAGACTAAGAAACACTGTGTCATCTGACCTGACCTGCGGACCTTCTCTGTG 399
Db 11 CysCysSerArgProLysValAla-----AlaLeuThrValGlyThrLeuLeu--- 26

QY 400 GAGCTGGCTGGCGCTGCTACTCTGGAAGTTTCATGGCAGCAAGTGTCTCAACTCT 459
Db 27 -----PheLeu----- 28
QY 460 GGGATAGTGGACTCTCTCAGGTACCTGCATCAACCCCTCTAACTGTGTGTGATGGCGTG 519
Db 29 -----ThGlyIleGlyAlaAspTrpAlaIleValThr 40
QY 520 TCACACTGCCCGCGGGAGGACGAGATCGGTGTGTTCGCTCTCAGGACCAAACTTC 579
Db 41 IleLeuLeuArgSerAspGlnGluProLeuTyrGlnValGlnLeuSerProGlyAspSer 60
QY 580 ATCTTTTCTGCTGACTCTCTCAGAGGAAGTCTCGGCACCTGTGTGTCAGACGACTGG 639
Db 61 ArgLeuLeuValLeuAspLysThrGluGlyThrTrpArgLeuLeuCysSerSerArgSer 80
QY 640 AACGAGAACTACGGCGCGGCTCGAGGACATGGCTATATAAGATAAATTTTACTCT 699
Db 81 AsnAlaArgValAlaGlyLeuGlyCysGluGluMetGlyPheLeuArgAlaLeuAlaHis 100
QY 700 AGCCAAGAAATAGTGGATGACAGCGATCCACGACTTTATGAAACTGAACACAGTGCC 759
Db 101 SerGluLeuAspValArgThrAlaGlyAlaAsn-----GlyThrSerGly 115
QY 760 -----GGCAATGTCTGATATCTATAAAACTGTACCAAGTATGCC 801
Db 116 PhePheCysValAspGluGlyLeuProLeuAlaGlnArgLeuLeuAspValIleSer 135
QY 802 -----TGTTCTTCAAAAGCAGTGGTTTCTTACGCTGTATAGCTTGGGGGTCAAC 852
Db 136 ValCysAspCysProArgGlyArgPheLeuThrAlaThrCysGlnAspCysGly----- 153
QY 853 TTGAACTCAAGCCCGCAGACAGATGTTGGCGCGGAGAGCGCGCTCCCGGGCGCTGG 912
Db 154 ArgArgLysLeuProValAspArgIleValGlyGlyGlnAspSerSerLeuGlyArgTrp 173
QY 913 CCCTGGGAGGTGACCTTCGACGCTCCAGACGTCACGCTGTGCGGAGGCTCCATCATCACC 972
Db 174 ProTrpGlnValSerLeuArgTyrAspGlyThrHisLeuCysGlyGlySerLeuLeuSer 193
QY 973 CCGAGTGGATCTGACAGCGCGCTGCTGGTGGAAAAAAGCTCTTAACTCAATCCATGCAT 1032
Db 194 GlyAspTrpValLeuThrAlaAlaHisCysPheProGluArgAsnArgValLeuSerArg 213
QY 1033 TGGACGGCATTTGGCGGATTTTGAGACAATCTTTCATGTCTTCTATGAGCGCGGATACCAA 1092
Db 214 TrpArgValPheAlaGlyAlaValAlaArgThrSerPro---HisAlaValGlnLeuGly 232
QY 1093 GTAGAAAAAGTATTTCTCTCCAAATTAT-----GACTCCAGACC 1134
Db 233 ValGlnAlaValIleTyrHisGlyGlyTyrLeuProPheArgAspProThrIleAspGlu 252
QY 1135 AAGAAACATGACATTTGGCTGATGAGCTGCAGAGCTCTGACTTTTCAACGACCTAGTG 1194
Db 253 AsnSerAsnAspIleAlaLeuValHisLeuSerSerLeuProLeuThrGluTyrIle 272
QY 1195 AAACCATGTGTCTGCCAACCCAGGATGATGCTGCGAGCAGACAGCTCTGCTGGAATT 1254
Db 273 GlnProValCysLeuProAlaAlaGlyGlnAlaLeuValAspGlyLysValCysThrVal 292
QY 1255 TCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAACTGTCTGAAAGCTGCCAAG 1314
Db 293 ThrGlyTrpGlyAsnThrGlnPheTyrGlyGlnGlnAlaValValLeuGlnGluAlaArg 312
QY 1315 GTGCTTCTCATGACACAGATGCAACAGCAGATATGCTATGATGCAACCTGATCACA 1374
Db 313 ValProIleIleSerAsnGluValCysAsnSerProAspPheTyrGlyAsnGlnIleLys 332
QY 1375 CCAGCCATGATCTGTGCGGCTTCTCGACGGGAAAGCTCGACTTTCGCCAGGTGACAGT 1434
Db 333 ProLysMetPheCysAlaGlyTyrProGluGlyGlyIleAspAlaCysGlnGlyAspSer 352

QY 1435 GGAGGGCTCTGTC-----ACTTCGAAGAACAAATATCTCGTGGCTGATAGG 1482
|||
Db 353 GlyGlyHisPheValCysGluAspArgIleSerGlyThrSerArgIlePheValCysGly 372
|||
QY 1483 GATACAGCTGGGTTCTCGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTG 1542
|||
Db 373 IleValSerTrpGlyThrGlyCysAlaLeuAlaArgLysProGlyValThrLysVal 392
|||
QY 1543 ATGTTATTCAGGACTGGATTTATCGACAATAGAGGCGACAGCGCTAATCCACATGGTCT 1602
|||
Db 393 IleAspPheArgGluTrpIlePhePheGlnAlaIleLysThr----- 405
|||
QY 1603 TCGTCCTTGACGTCTTTACAGAAACAAATCGGGCTGGTTCCTTCCTCC 1653
|||
Db 406 -----HisSerGluAlaThrGlyMetValThrGlnPro 416
|||
RESULT 13
T30337
polyprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
C:Accession: T30337
R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from xeno-
laevis oocytes.
A:Reference number: 220829
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <YAN>
A:Cross-references: EMBL:U81290; NID:g2981640; PID:g2981641; PIDN:AAC24717.1
C:Superfamily: trypsin related polyprotein; trypsin homology
Alignment Scores:
Pred. No.: 2.1e-28 Length: 1524
Score: 504.50 Matches: 132
Percent Similarity: 46.46% Conservative: 65
Best Local Similarity: 31.13% Mismatches: 140
Query Match: 15.87% Indels: 87
DB: 2 Gaps: 14
US-09-323-597c-1 (1-1738) x T30337 (1-1524)
QY 511 GATGGCTGTACACACTCCCGGGGAGAGAGAGAGATCGTGTGTCCTCTACGGA 570
|||
Db 486 AsplaValGluValThrAspGlyValGluGluLysGlnLeuIleAlaArgLeuValCysGly 505
|||
QY 571 CCAAACTTCATCCTCAGGTGTACTCATCTCAGAGGAGTCTCGGACCCCTGTGTGCCAA 630
|||
Db 506 TyrThrLeuProLeuProIle--SerSerProGluAsnThrMetLeuIleArgPheLys 524
|||
QY 631 GACACTGGACGAGAACTACGGCGGGCGGCTCGAGGACATGGGCTATAGAAATAT 690
|||
Db 525 ThrAspMetGluAsnSerTyr-----ProGlyPheLysValLys 537
|||
QY 691 TTT-----TACTCTACCAAGAAATAGTGGATGACAGCGGA 726
|||
Db 538 PheSerPheValProLysGluLysGlnPheSerLeuPro-----ValAspAspThrPro 555
|||
QY 727 TCCACAGCTTTATGAAGAACTGAACAGAGTCCGGGCAATGCGATATCTATATAAACTG 786
|||
Db 556 ThrIleSerMet-----LeuHisProArgAlaIleAlaLeuAspVal----- 569
|||
QY 787 TACCACAGTGTATGCTGTCTTCAAAAGCAGTGTGTTCTTACGCTGTATAGCTGCGGG 846
|||
Db 570 -----CysGly 571
|||
QY 847 GTCACACTGTAACCTCAACGCGC-----CAGAGCAGGATGTGGCGCGAGAGCGGCTC 900
|||
Db 572 MetAlaProMetThrProLysTrpTrpLeuProArgIleValGlyGlyGluGluAlaSer 591
|||
QY 901 CCGGGGCGCTGGCGCTGCGAGTCCAGCTCCAGAACAGTCCACAGTGTGCGGAGGC 960
|||

Db 592 ProAsnSerTrpProTrpGlnValGlnIlePhePheLeuArgThrPheHisCysGluGly 611
|||
QY 961 TCCATCATCACCCCGAGTGGATCTGACAGAGCGCCACTCGTGGGAAAAACCTCTTAAC 1020
|||
Db 612 AlaIleIleSerProGlnTrpIleLeuThrAlaAlaHisCysIleArgAlaAla----- 629
|||
QY 1021 AATCCATGTCATGGACGGCATTTGGGGGATTTTGAGACAATCTTTTCATGTCTATGGA 1080
|||
Db 630 GluProSerTyrTrpThrValIleAlaGly-----AspHisAsnArgMetLeuAsnGlu 647
|||
QY 1081 GCGGGATACAGTAGAGAAAAGTGATTTCT-----CATCAAATATGATCTCAAG 1131
|||
Db 648 SerThrGluGlnIleArgAsnIleThrIleArgIleHisAspAsnTyrAsnSerGlu 667
|||
QY 1132 ACCAAGACAAATGACATTTGGCTGATGAAGCTGCAGAAGCTCTGATTTTCAACGACCTA 1191
|||
Db 668 ThrTyrAspAsnAspIleAlaLeuLeuTyrLeuGluGluProLeuAspLeuAsnAspPhe 687
|||
QY 1192 GTGAACCGAGTGTCTGCCAACCCAGGATGATCTGCAGCAGCAGACAGCTCTGCTGG 1251
|||
Db 688 ValArgProValCysLeuProGluGluValLeuThrProAlaSerValCysVal 707
|||
QY 1252 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTCTGAACGCTGCC 1311
|||
Db 708 ValThrGlyTrpGlyAsnThrAlaGluAspGlyGlnProAlaLeuGlyLeuGlnLeu 727
|||
QY 1312 AAGTGCTCTTCTCATTCAGACACAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
|||
Db 728 GlnLeuProIleLeuAspSerIleIleCysAsnThrSerTyr---TyrSerGlyGluLeu 746
|||
QY 1372 ACACAGCAGCATGATCTGCGCGCTCTCTGCAGGGG---AACGTCGATTTCTGCCAGGT 1428
|||
Db 747 ThrAspHisMetLeuCysAlaGlyPheProSerSerLysGluLysAspAlaCysGly 766
|||
QY 1429 GACAGTGGAGGGCTCTGCTC---ACTTCGAAGAACAAATATCTGTCGGCTGATAGGGA 1485
|||
Db 767 AspSerGlyGlyProLeuValCysGlnAsnGluLysGlnPheSerIleTyrGlyLeu 786
|||
QY 1486 ACAAGTGGGGTCTCGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGATG 1545
|||
Db 787 ValSerTrpGlyGlyCysGlyArgValSerLysProGlyValThrLysValArg 806
|||
QY 1546 GTATTCACGACTGGATTTATCGACAATAGAGGCGAGCGCTAATCCACATGCTCTTCG 1605
|||
Db 807 LeuPhePheThrTrpIleGln----- 813
|||
QY 1606 TCCTTGACGTGCTTTTACAGAAACAAATGCGGCTCGTTCCTCCCGTCATGATTT 1665
|||
Db 813 ----- 813
|||
QY 1666 ACTCTTAGAGATGATTCAGAGGTCACTTCATTTTATAAAGTGAACCTGTGTGCGCAA 1725
|||
Db 814 AsnThrGlnGlnAspLeuGlnGlnGluAsnAlaLeuAsnSerLysSerValGluGln 833
|||
QY 1726 AAAAAA 1737
|||
Db 834 ArgGluGlyLys 837
|||
RESULT 14
S11674
acrosin (EC 3.4.21.10) precursor - human
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S11674; S23499; S12063; A61022; S03330
R:Keime, S.; Adham, I.M.; Engel, W.
Eur. J. Biochem. 190, 195-200, 1990
A:Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene
A:Reference number: S11674; MUID:90306003; PMID:2114285
A:Accession: S11674
A:Molecule type: DNA
A:Residues: 1-421 <KEI>
A:Cross-references: EMBL:X54017; NID:G35582; PIDN:CAA37964.1; PID:G1216165
A>Note: the authors translated the codon AGG for residue 64 as Thr and CTC for residue

940 QY AAGCTCCAGTGTGGAGAGCTCCATCATCATCCCCGAGTGGATCTGTGACAGCGCGCCAC 999

Db 69 A-gTyrHisThrCysGlyGlySerLeuLeuAAsnSerArgTrpValIleThrAlaAlaHis 88

QY 1000 TGC---GTGGAAAACCTCTTAACAATCCATGCATTCGACGCGCAATTCGGCGGAATTTG 1056

Db 89 CysPheValGlyLys-----AsnAsnValHisAspTrpArgLeuValPheGlyAlaLys 106

QY 1057 AGACAATCTTTCATGTTCTATGGA-----GCCGGATACCAA----- 1092

Db 107 Glu-----IleThrTyrGlyAsnAsnLysProValIysAlaProLeuGlnGluArg 123

QY 1093 ---GTAGAAAAGTATTCTTCATCCAAATTATGACTCCAAAGACCACGAACAATGACATT 1149

Db 124 TyrValGlnLysIleIleHisGlnLysTyrAsnSerAlaThrGlnGlyAsnAspIle 143

QY 1150 GGCTGATGAGCTGCAGAGGCTCTGACTTTTCAACGACTAGTGAACACAGTGTGTCTG 1209

Db 144 AlaLeuValGluIleThrProPheIleSerCysGlyArgPheIleGlyProGlyCysLeu 163

QY 1210 CCACAC---CCAGGCATGATGCTGCAGCCAGAACACAGCTCTGCTGGGATTTCCGGAGTGGGG 1266

Db 164 ProHisPheLysAlaGlyLeuProArgGlySerGlnSerCysThrValAlaGlyTrpGly 183

QY 1267 GCCACCGAGGAGAAAGGG---AAGACTCTCAGAAGTCTCTGCAACCTCCCAAGTGCTTCTC 1323

Db 184 TyrIleGlnGluLysAlaProArgProSerSerIleLeuMetGluAlaArgValAspLeu 203

QY 1324 ATTGAGACACAGAGATGCAACAGCACATATGCTATGACACACCTGATCATCACAGCCCATG 1383

Db 204 IleAspLeuAspLeuCysAsnSerThrGlnTrpTyrAsnGlyArgValGlnProThrAsn 223

QY 1384 ATCTGTGCCGGCTTCCTGCAGGGGAAACGTCAATTCTTCAGGGGTGACAGTGGAGGCGCT 1443

Db 224 ValCysAlaGlyTyrProValGlyLysIleAspThrCysGlnGlyAspSerGlyGlyPro 243

QY 1444 CTGGTC-----ACTTCGAAGAACATATCTGTGTGCTGATAGGGGATACAGCTGGGGT 1497

Db 244 LeuMetCysLysAspSerLysGluSerAlaTyrValValGlyIleThrSerTrpGly 263

QY 1498 TCTGGCTGTGCCAAACCTTACAGACCAGGAGTGTACGGGAATGTGTGGTATTTCACGGAC 1557

Db 264 ValGlyCysAlaLeuAlaLysArgProGlyIleTyrThrAlaThrTrpProTyrLeuAsn 283

QY 1558 TGGATT 1563

Db 284 TrpIle 285

RESULT 15

JX0172

acrosin (BC 3.4.21.10) precursor form 3 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000

C:Accession: JX0172; JX0138

R:Watanabe, K.; Baba, T.; Kashiwabara, S.; Okamoto, A.; Arai, Y.

J. Biochem. 109, 828-833, 1991

A:Title: Structure and organization of the mouse acrosin gene.

A:Reference number: JX0172; MUID:92041732; PMID:1939002

A:Accession: JX0172

A:Molecule type: DNA

A:Residues: 1-436 <WAT>

A:Cross-references: GB:S66245; NID:G938706; PIDN:AAB20293.1; PID:G238707

R:Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.

J. Biochem. 108, 785-791, 1990

A:Title: Primary structure of mouse proacrosin deduced from the cDNA sequence

A:Reference number: JX0138; MUID:91185335; PMID:2127931

A:Accession: JX0138

A:Molecule type: mRNA

A:Residues: 4-436 <KAS>

A:Cross-references: GB:D00754; NID:G220322; PIDN:BAA00651.1; PID:G220323

C:Comment: Acrosin is an acrosomal protease that plays an important role in the e-binding activity.

C:Genetics:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 15:07:22 ; Search time 91 Seconds
(without alignments)
11890.358 Million cell updates/sec

Title: US-09-323-597C-1

Perfect score: 3179

Sequence: 1 ggcggagcgagcgagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Zgapop 6.0 , Zgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 2553080

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09323597/runat_06072004_111730_29788/app_query.fasta_1.1927
-DB=Published Applications AA -QFWT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=hblosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09323597@cgn 1 1 13 @runat_06072004_111730_29788
-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description

1	2717	85.5	492	15	US-10-295-027-586	Sequence 586, App
2	2711	85.3	492	14	US-10-205-823-415	Sequence 415, App
3	2709	85.2	492	9	US-09-759-143-932	Sequence 932, App
4	2709	85.2	492	9	US-09-780-669-932	Sequence 932, App
5	2709	85.2	492	9	US-09-822-827-932	Sequence 932, App
6	2709	85.2	492	9	US-09-895-793-932	Sequence 932, App
7	2709	85.2	492	9	US-09-895-814-932	Sequence 932, App
8	2709	85.2	492	13	US-10-012-896-932	Sequence 932, App
9	2709	85.2	492	14	US-10-144-678A-932	Sequence 932, App
10	2709	85.2	492	14	US-10-294-025-932	Sequence 932, App
11	2704	85.1	492	14	US-10-334-038-6	Sequence 6, Appl
12	2701	85.0	492	12	US-09-825-751A-63	Sequence 63, Appl
13	2696	84.8	492	9	US-09-759-143-895	Sequence 895, App
14	2696	84.8	492	9	US-09-780-669-895	Sequence 895, App
15	2696	84.8	492	9	US-09-879-793-14	Sequence 14, Appl
16	2696	84.8	492	9	US-09-822-827-895	Sequence 895, App
17	2696	84.8	492	9	US-09-895-793-895	Sequence 895, App
18	2696	84.8	492	9	US-09-895-814-895	Sequence 895, App
19	2696	84.8	492	10	US-09-776-191-70	Sequence 70, Appl
20	2696	84.8	492	13	US-10-012-896-895	Sequence 895, App
21	2696	84.8	492	14	US-10-144-678A-895	Sequence 895, App
22	2696	84.8	492	14	US-10-294-025-895	Sequence 895, App
23	2696	84.8	492	15	US-10-295-027-588	Sequence 588, App
24	2696	84.8	492	15	US-10-295-027-871	Sequence 871, App
25	2696	84.8	492	15	US-10-158-214A-37	Sequence 37, Appl
26	2696	84.8	492	16	US-10-188-832-204	Sequence 204, Appl
27	2171	68.3	393	9	US-09-759-143-934	Sequence 934, App
28	2171	68.3	393	9	US-09-780-669-934	Sequence 934, App
29	2171	68.3	393	9	US-09-822-827-934	Sequence 934, App
30	2171	68.3	393	9	US-09-895-793-934	Sequence 934, App
31	2171	68.3	393	9	US-09-895-814-934	Sequence 934, App
32	2171	68.3	393	13	US-10-012-896-934	Sequence 934, App
33	2171	68.3	393	14	US-10-144-678A-934	Sequence 934, App
34	2171	68.3	393	14	US-10-294-025-934	Sequence 934, App
35	2126	66.9	384	9	US-09-981-353-23	Sequence 23, Appl
36	1540	48.4	283	9	US-09-988-975A-1	Sequence 1, Appl
37	1263	39.7	229	14	US-10-235-699-1	Sequence 53, Appl
38	1165	36.6	209	9	US-09-759-143-897	Sequence 897, App
39	1165	36.6	209	9	US-09-780-669-897	Sequence 897, App
40	1165	36.6	209	9	US-09-822-827-897	Sequence 897, App
41	1165	36.6	209	9	US-09-895-793-897	Sequence 897, App
42	1165	36.6	209	9	US-09-895-814-897	Sequence 897, App
43	1165	36.6	209	13	US-10-012-896-897	Sequence 897, App
44	1165	36.6	209	14	US-10-144-678A-897	Sequence 897, App
45	1165	36.6	209	14	US-10-144-678A-897	Sequence 897, App

ALIGNMENTS

RESULT 1
US-10-295-027-586
; Sequence 586, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevizi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Ros Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 586
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-295-027-586

Alignment Scores:

Pred. No.: 4,28e-207 Length: 492
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.47% Indels: 0
 DB: 15 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-295-027-586 (1-492)

QY 112 ATGGCTTTCACTACGGGTACACAGCTATTGGACCTTACTATGAAACCATGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTy-TyrGluAsnHisGlyTyr 20
 QY 172 CAACGGAAACCCCTATCCGACAGCCCATCTGGTCCCACTGTCTACAGGTGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 QY 232 CGGGCTCAGTACTACCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCC 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCCGTCTGTCACGACGCCCAATCCCATCCGGACAGTGTGCACTCAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAGAAAGCACTGTCATCACTTGACCTGGGACCTTCTCTGGGAGCTCGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 QY 412 GCGCTGGCTACTCTGGAAGTTTCATGGCCCAAGTGTCCAACTCTGGGATAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCAGGTACCTGCATCAACCCCTTAACCTGTGATGGGTGTCACTGCCCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 QY 532 GCGGGGAGGACGAGAAATCGGTGTTCGCTCTACGGACCAAACTTCTATCTTCAGGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 QY 592 TACTATCTCAGAGGAGTCTGGCACCTGTGTGCCAAGACACTGGAACGAGAACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTyrPheIleProValCysGlnAspAspTyrAsnGluAsnTyr 180
 QY 652 GCGCGGGCGCTCGAGGACACTGGCTATAGATAATTTTACTCTAGCGCAAGATA 711

Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACTGAACAACAGTCCGGCAATGTCGAT 771
 Db 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACACAGTGTGCTTCTTCAAAAGCAGTGGTTCCTTTCACGC 831
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTCGCGGGTCAACTTGAACCTCAAGCCGACAGAGAGGATTGTGGCGGCGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGGCTCCCGGGGCTGGCCCTGGCAGTTCAGCTGACGTCACGTCGACAGCTCCAGAGCTCA 951
 Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TGGGAGGCTCCATCATCACCCCGAGTGTGAGTCGTGACAGCGCCCACTGCTGGTGA 1011
 Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTACAAATCCATGCGCATTGACGCGCATTTGGGGGATTTTCAGACATCTTTCATG 1071
 Db 301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgLysSerPheMet 320
 QY 1072 TTCTATGAGCGCGGATACCAAGTAGAAAACTGATTTCTCTCAAAATTTATGACTCCAA 1131
 Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 QY 1132 ACCAAGAACATGATTCGCTGATGAGCTGACAGAGCTGACAGCTGCTTCAAGACCTTA 1191
 Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 QY 1192 GTGAAACCCAGTGTCTGCTGCCCAACCCAGGATGATGTGCGACCCAGACACACTCTGCTGG 1251
 Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 QY 1252 ATTTCCGGGTGGGGGGCCACCGAGGAAAGAGGAGACCTCAGAGTGTCTGAACGCTGCC 1311
 Db 381 IleSerGlyTyrGlyAlaThrGluGluGlyLysThrSerGluValLeuAspAlaAla 400
 QY 1312 AAGGTGCTTCTCATGTAGACACAGATGCAACAGCAGATATGTCTATGACAACTGATC 1371
 Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 QY 1372 ACACAGCAGATGATGTGCGGCTTCTGCGAGGAAACGTCGATTTCTTGCAGGGTGAC 1431
 Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTCGAGGCGCTCTGGTCACTTCGAAACAATAATCTGGTGGCTGATAGGGGATACAGC 1491
 Db 441 SerGlyGlyProLeuValThrSerLysAsnIleTyrTrpLeuIleGlyAspThrSer 460
 QY 1492 TGGGGTTCGTGCTGCAAGTTCACAGACCGAGAGTGTACGGGATGTGATGGTATTC 1551
 Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 QY 1552 ACCGACTCGATTTATCGCAAAATGAGGCGCAGCGC 1587
 Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2

US-10-205-023-415

; Sequence 415, Application US/10205823

; Publication No. US20030108963A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Monahan, John E.

; APPLICANT: Endege, Wilson O.

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Gorbacheva, Bella

; APPLICANT: Hoersch, Sebastian

APPLICANT: Kamathkar, Shubhangi
APPLICANT: Wonsey, Angela M.
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Anderson, Dustin
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF PROSTATE CANCER
FILE REFERENCE: MRI-044
CURRENT APPLICATION NUMBER: US/10/205,823
CURRENT FILING DATE: 2002-07-25
PRIORITY APPLICATION NUMBER: 60/307,982
PRIORITY FILING DATE: 2001-07-25
PRIORITY APPLICATION NUMBER: 60/314,356
PRIORITY FILING DATE: 2001-08-22
PRIORITY APPLICATION NUMBER: 60/325,020
PRIORITY FILING DATE: 2001-09-25
PRIORITY APPLICATION NUMBER: 60/341,746
PRIORITY FILING DATE: 2001-12-12
PRIORITY APPLICATION NUMBER: 60/362,158
PRIORITY FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 415
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-10-205-823-415

Alignment Scores:
Pred. No.: 1,28e-206 Length: 492
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 85.28% Indels: 0
DB: 14 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-205-823-415 (1-492)

QY 112 ATGGCTTTGAACCTCAGGCTACCAACAGCTATGACCTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGGACAGCCCACTGCTGCTGCCACTGCTACAGGTCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGACTCAGTACTACCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCC 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTCGACGACGCCCAATCCCGTCCCGTCCCGTCCCGTCCCGTCC 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACCCTGACCTGACCTGGGACCTTCTCTGGGAGCTGCCGTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCCGCTGGCTTACTCTGGAAGTTTATGGGACGAGTCTCCAACTCTGGGATAGAGTGC 471
DB 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGGTACTGTCATCAACCCCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGAGGACGAGATCGGTGTGTTGCTGCTTACGACCAAACTTCTCATCTTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGAAAGTCTGCGACCCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 651

DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTTCAGGGACATGGGCTATAAGATAATTTTACTCTACCCAAAGGAATA 711
DB 181 GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGGATCCACCACTTATGAACCTGAACAACAGTCCGCAATGTGCAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTTACACACAGTGCCTGTCTTCAAAAACAGTGGTTCTTCTTACGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTTCGGGCTCACTTGAACCTGAGCGGCGGACAGAGGATGTGGGGGGGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
QY 892 AGCGCGCTCCGGGGGCTGCGCCCTGCGAGTGCAGCTGCACCTCCACAGACCTCCAGCTG 951
DB 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGGAGGCTCCATCATCACCCCGAGTGTGATGACAGCGGCGGCTCCGCTGGGAAAAA 1011
DB 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGCGCATTCGACGGCATTTTCGGGGATTTTGACAACTCTTTCATG 1071
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGAGCCCGGATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATTATGACTCCAAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTCGGCTGATGAGAGCTGCAGAGGCTCTGACTTTCACAGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCCAGTGTGTCTGCCCAACCCAGGCAATGCTGCAGCGCAGAACAGCTCTCTCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAGTGTGTAAGCGTGC 1311
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGCTGCTTCTCATTCAGACACAGATGCAACAGCAGATATCTCTATGACAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCGACCATGATCTGTCGGCTTCTGCGAGGAAAGTCTGATTCCTTCCAGGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGGCTCTGTGCTCACTTCGAAGAACAATATCTGCTGGCTGATAGGGGATACAAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TCGGGTCTGCTGCTGCCAAAGCTTACAGACAGGAGTGTACCGGAATGTGATGTGTTTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGATTTTATCGACAAATGAGGCGACAGCGC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 3

US-09-759-143-932
; Sequence 932, Application US/09759143
; Patent No. US2002002248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.


```
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Alijun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2003-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-759-143-932 (1-492)

QY 112 ATGGCTTTGAATCAAGGGTACACACAGCTATTGACCTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACGGGAAACCCCTATCCCGCAGAGCCACTGTGGTCCCACTGTCTACAGATGCAAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CGGGCTCAGTACTACCGTCCCGCCCGCCAGTACGCGCCGAGGCTCTGACGCGAGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGCTCTGCAAGCAGCCCAATCCCACTCCGGACAGTGTGACACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAGCACTGTGATCACCCTGACCTGACCTGGGAGCTTCTGCTGGAGCTCGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCCTACTCTGGAAGTTCATGGGCGACAAAGTGTCTCAACTCTGGGATAGATGC 471
DB 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACCTGATCAACCCCTCACTGCTGATGCTGATGGCTGTACATGCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGAAATCGGTGTCTGCTCTACGACCAAACTTCACTCTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleGluVal 160
QY 592 TACTCATCTCAGAGAAATCTCGGCACTGTGTGCGCACTGTGTGCGCACTGCAAGACGACTG 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
```

RESULT 4

```
US-09-780-669-932
; Sequence 932, Application US/09780669
; Patent No. US2002005197A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
```

```
QY 652 GGGGGGGCGGCTCGAGGGACATGGGCTTATAAGATAATTTTACTCTAGCAAGGAATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAAATGAAACACAAAGTGGCGGCAATGTCCAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTATACACAGTGTCCCTGTTCTTCAAAAGCAGTGGTTTCTTACGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTTCGGGGTCACTTGAACCTCAAGCCCGCAGACGAGATTGTGGCGGGCAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGACCTGACGCTCCAGACGCTCCACGTG 951
DB 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCCATCATCACCCCGAGTGGATCTGACAGCGCCCACTGCTGGGAAAAA 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACATCCATGSCATTTGACGCGCATTTGGGGGGATTTTGAGACAACTCTTTCATG 1071
DB 301 ProLeuAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGGATACCAAGTAGAAAAAGTGTATTTCTCATCCAAATTTATGATCCCAAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTTGCTGATGACGCTGACAGAGCTCTGACTTTTCAGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACCAAGTGTCTGCGCCCAACCCAGGATGATGCTGACGCGCAGAACAGCTCTGCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGGCCACGAGGAGAAAGGAGAACCTCAGAGTGTGGAAGCTGCC 1311
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGTCTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACCAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCAACCATGATCTGTGCGGCTTCTGCGGGGAACTGCGATTCTTGCAGGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGGCTCTGGTCACTTCGAAACAACAATATCTGGTGGCTGATAGGGGATACAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
QY 1492 TGGGGTTCTGGCTGTGCAAAAGCTTACAGACCCAGGAGTGTACGGGAATGTGATGATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACCGACTGGATTATCGAATAATCAGGGCAGAGCGC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
```

APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedwick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780,669
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-780-669-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Idels: 0
DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-780-669-932 (1-492)

Qy 112 ATGCTTTGAAGTCAAGGTCACCAAGTATTCGACCTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20

Qy 172 CAACCGGMAAACCCCTATCCCGGACAGCCACTGTGTGTCCTGCTACGAGGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40

Qy 232 CCGGCTCAGTACTACCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCC 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProAlaProAlaGlnAla 60

Qy 292 TCCAAACCCCGTCTGTGACAGCAGCCAAATCCCAATCCCGGACAGTGTGCACCTCAA 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

Qy 352 ACTAAGAAAGCAGTGTGCATCAGCTTACCTGGGACCTTCTCTGGGAGCTGCGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100

Qy 412 GCGCTGCTCTCTGGAAGTTATGCGGACAGCAAGTCTCCAACTCTGGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheValGlySerLysCysSerAsnSerGlyIleGluCys 120

Qy 472 GACTCTCTCAGTGTGCATCAGCTTACCTGGGACCTTCTCTGGGAGCTGCGCTG 531
Db 121 AspSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140

Qy 532 GCGGGGAGGACGAGAAATCGGTGTGTTCGCTCTTACCGACCAAACTTCTCTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160

Qy 592 TACTCATCTCAGAGGAAGTCTCTGCGCACCTGTGTGTGCCAAGACGACTGGAAACGAACTAC 651

Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
Qy 652 GGGCGGCGGCTGTGACGGGACATGGGCTATAAGAAATAATTTTACTCTAGCAGGAAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Qy 712 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACAAGTGGCGGCAATGTGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Qy 772 ATCTATAAAAAAAGTGTACACAGTGTCTTCTTCAAAAGCAGTGGTTCCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
Qy 832 TGTATAGCTTGGCGGTCACACTTCAACTAAGCCCGCAGACGAGATGTGGCGGCGGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Qy 892 AGCGCGCTCCCGGGGCGCTGCGCTGCGAGTGTGACAGCTGACAGTCCAGAGACGTCCAGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Qy 952 TGGGAGGCTCCATCATCATCACCCCGAGTGTGTGACACCCCGCCACTGCGTGGAAAAA 1011
Db 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Qy 1012 CCTCTTAACAATCCATGGCATTGGAGCGCATTTGGGGGATTTTGACACAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

Qy 1072 TTCTATGAGCCCGGATACCAAGTAGAAAAAGTATTTCTCATCAAATATGACTTCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340

Qy 1132 ACCAAGAACCAATCAGATTTGGCTGTGATGACAGCTGACAGAGCTCTGACTTTCAACGACTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360

Qy 1192 GTGAAACCCAGTGTGTCTGCCAACCCAGGCGATCATGTGACAGCCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380

Qy 1252 ATTTCCGGTGGGGGCGCCACCGAGGAGAAAGGAAAGAGCTCTGAGAGTGTGACGCTGCC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400

Qy 1312 AAGTGTCTTCTCATTGAGACACAGAGATGCAACAGCAGATATCTCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420

Qy 1372 ACACGAGCCATGATCTGTGCGGCTTCTGCGAGGGAACGTTCGATTTCTGCCAGGCTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

Qy 1432 AGTGGAGGCGCTCTGCTGCTCGAAGAACATATCTGCTGGCTGATAGGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460

Qy 1492 TGGGGTCTCTGCTGTGCCAAAGCTTACAGACGAGGAGTGTACGGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480

Qy 1552 ACGGACTGATTTATCGACAAATAGGCGGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 5
US-09-822-827-932
; Sequence 932, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.534C1
 ; CURRENT APPLICATION NUMBER: US/09/822,827
 ; CURRENT FILING DATE: 2001-03-28
 ; NUMBER OF SEQ ID NOS: 982
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 932
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-822-827-932

Alignment Scores:
 Pred. No.: 1,85e-206 Length: 492
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 85.22% Indels: 0
 DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-822-827-932 (1-492)

```

QY 112 ATGCTTTGAACCTCAGGTGACCCAGCTATTCGACCTTACTATGAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaLeuGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCCACAGCCCACTGTGTCCCACTGTCTACGAGGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGCTCAGTACTACCGTCCCGTCCCGCCAGTACGCCGAGGGTCTTGACGAGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTCAGCAGCCCAATCCCATCCCGGAGAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAGCACTGTGTATCACCCTTGACCCCTGGGACCTTCTCGTGGAGTGCCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCCCTGCGCTACTCTGGAAGTTCATGGGAGCAGTCTCCAACTCTGGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACTCCTCAACCCCTCTAACTCGTGTGTATGGCGTGTCCACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCGGTGTGGCTCTACCGACCAACTTCTCCTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAAGTCTCGCACCTGTGTGCCAAGACCACTGGAAACGAGAACTAC 651
Db 161 TyrSerSerGlnA-glySerSerTyrHisProValCysGlnAspSerTyrAsnGluAsnTyr 180
QY 652 GGGGGGGGGCTCAGGACATGGGTATAGAAATATTTTCTACTTACCCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACACGGGATCCACAGCTTTATGAACCTGAACCAAGTCCCGCAATGTGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACCAAGTATGCTGTCTTCAAAAGCAGTGGTTCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGCGGTCAACTCAACTCAAGCGCCAGAGGATGTGGCGGCGAG 891

```

```

Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCTGGCCCTCGCAGGTGACGCTGCACGTCCAGAACCTCCAGGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGCGAGGTCCCATCATCATCCCGCGAGTGTGATCGTGCACAGCCGCCCACTCGGTGAAA 1011
Db 281 CysGlyLysSerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGCATTTGGACGGCAATTTGGCGGGATTTTGACACAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGAGCGGTACCAAGTAGAAAAGTAGATTCTCATCAATATATGACTCCAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAATGACATTCGCTGATGAAGCTGCAGAAGCCTCTGACTTTTCAACGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACCAAGTGTCTGCTGCCCAACCCAGGATGATCTCTGAGCCAGCAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
QY 1252 ATTTCCGGGTGGGGGCCCAACCGAGGAGAAAGGAGAACCTCAGAGTGTCTGAAACGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTCTCTCTCATTTGAGACACAGATGCAACAGAGATATCTTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCAACCATGATCTGTGCGCTCTCTCGAGGGAACGTTCGATCTTCCAGGCTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTGCTCCTTCTGAAAGAACATATCTGTGTGCTGTAGGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGCTGTGCCAAAGCTTACAGACAGAGTGTACGGGAATCTGATGTTGTTTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGATTATTCGACAAATGAGCGCAGACGCC 1587
Db 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492

```

RESULT 6

US-09-895-793-932
 ; Sequence 932, Application US/09895793
 ; Publication No. US20020192763A1
 ; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuqiu
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.

```

; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-932

Alignment Scores:
Pred. No.: 1.85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-895-793-932 (1-492)
QY 112 ATGGCTTCAACTCAGGGTCACACAGCTATTGCACTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCAGCCACTGTGTCGCCACTGTGTCCCACTGTCTACGAGGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CGGCTCAGTACTACCGCTCCCGCTGCCCGCCAGTACGCCCGAGGGTCTGTGACGAGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCACACCCGCTGTCTGCACGACGAGCCCAATCCCACTCCGGGACAGTGTGCACCTCAAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAGAAAGACTGTGCATCACTTGNACCTTGNACCTCTGGGACCTTCTCTGTGGAGCTGCGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTACTCTCGAAGTTTCATGCGACGACAGTCTCCCACTCTGGGATAGAGTGC 471
DB 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTCTAATCTGGTGTGATGCGGTGCACACTGCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGACGAGAGAACTGGTCTGCTTGGCTCTTACGACCAAACTTCATCTCTTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAAGTCTCGGACCCCTGTGTGCCAAGACGACTGGAAACGAGACTAC 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGCGGGCGGCTCGGGACATGGGCTATTAAGAAATAATTTTACTCTAGCCCAAGNATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACAGCTTTATGAAACTGAACACAAGTGGCGGCAATGTTCGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAAAAGCTGTACACAGTATCCCTGTCTTCTTCAAAGCAGTGGTTTCTTTACGC 831

```

APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanser, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C26
CURRENT APPLICATION NUMBER: US/09/895,814
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 990
SOFTWARE: PastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-814-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 9 Gaps: 0

US-09-323-597c-1 (1-1738) x US-09-895-814-932 (1-492)

QY 112 ATGGCTTTGAATCAAGGTCACACAGACTATGGACCTTACTATGAACCAATGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACGGGAAACCCCTATCCCGACAGCCACCTGCTGCTCCACCTGCTAGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CGGCTCACTACTACCCGTCCTCCCGTGGCCCAAGTACGCCCGAGGCTCTGACGAGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCAACCCCGTCTGTGACGACAGCCCAATCCCATCCGACAGTGTGACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGGACCTTCTCTGTGGAGCTCGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCTGGCCTACTCTGGAAGTTTCATGGSCAGCAAGTGTCCAACTCTGGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTCACTGCTGTGTGATGGGTGTCACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATGGTGTCTGCTCTACGACCAAACTTCATCTTCAGTGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAGCTCTGGCACCTGTGTGCCAGACGACTGGAACGAGACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
QY 652 GCGCGGGGCGCTGACGAGGACATGGGCTATAGAAATAATTTTACTTAGCCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACGGGATCCACGAGCTTTATGAACCTGAACAGAGTCCCGGCAATGTCCAT 771

Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTATCCACAGTATGCTCTTCTTCAAAAGCAGTGGTTCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTGCGGGTCAACTTCAACTCAAGCCGACAGCAGCAGATTTGTGGGGCGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGly 260
QY 892 AGCGCGCTCCCGGGGCTGCGCCCTGCGAGTCAAGCTGCGACGCTCCAGAACGTCACG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGCGAGGCTCCATCATCACCCCGAGTGTGTGACGCGCCGCTGCTGCTGCTGCTG 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGCGCATTTGACGCGCATTTTGGGGGATTTTGAGACATCTTT 1071
Db 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGGATACCAAGTAGAAAAAGTGTCTTCTCATCCAAATTATGATCC 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACCAATGACATTGCGCTGATCAAGCTGACAGAGCTCTGCTTCAACGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACAGTGTCTGCCCCAACCCAGGCAATGATGCTGACGCCAGACAGCTCTGCTG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGSGGGGGCCACCGAGAGAAAGGAGAGCTCAGAAAGTGTCTGAAAGCTGCC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGCTTCTCTATTGAGACACAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCGCCATGATCTGCGCGCTCTCTGCGGGAGACGTCGATTTCTGCCAGGTCAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGGCTCTGCTGCTCACTTCGAGAACAAATATCTGCTGGCTGTAGGGGATAC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGGCTGCGCAAGCTTACAGACCGAGGTGTACGGGAATGTGATCGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGATTATTCGACAAATGAGGGCGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 8
US-10-012-896-932
Sequence 932, Application US/10012896
Publication No. US20020183251A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuguo
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.

```

; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Mesgher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012.896
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-932

Alignment Scores:
Pred. No.: 1.85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 13 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-012-896-932 (1-492)

Qy 112 ATGCTTGAAGTACAGGTGACAGGCTTGGACCTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrGluAsnHisGlyTyr 20

Qy 172 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCACTGTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40

Qy 232 CCGGCTCAGTACTACCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60

Qy 292 TCCAAACCCGTCGTGACAGCAGCCCAATCCCAATCCCGTCCCGTCCCGTCCCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

Qy 352 ACTAAGAAGCACTGTGCATCACCTTGACCTGGGACCTTCCTCGTGGAGCTGCGCTG 411
Db 81 ThrLysLysAlaLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100

Qy 412 GCGGCTGCTTACTCTGGAAGTTCATGGCAGCAGTCTCCCACTCTGGGATAGTGC 471
Db 101 AlaAlaGlyLeuLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120

Qy 472 GACTCCTCAGTACTGCATCAACCTCTCACTCACTCACTCACTCACTCACTCACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140

Qy 532 GCGGGGAGGACGAGATCGGTGTGTCCTCTACGACCAACTTCCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160

Qy 592 TACTCATCTCAGGAAGTCTGCGCACCTGTGTGTCGCAAGACGACTCGAAGCAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTrpAsnGluAsnTyr 180

```

```

Qy 652 GGGCGGGCGGCGCTGACAGGACATGGGCTATAAGATAATATTTTACTTCTAGCCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200

Qy 712 GTGATGACAGCGGATCCACAGCTTTATGAACACTGAACACAAAGTCCGGCAATGTGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220

Qy 772 ATCTATAAAAACTGTACACACAGTGTGCTGTCTTCAAAAGCAGTGGTTCCTTACGC 831
Db 221 IleTyrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 240

Qy 832 TGTATAGCTTGGCGGTCACACTTGAACCTCAAGCCGACAGAGAGTGTGGCGGCGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260

Qy 892 AGCGCGCTCCCGGGGCGCTGGCCCTGCGCAGTGTGACCTGCACCTCCAGAACCTCCAGTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280

Qy 952 TGGGAGGCTCCATCATCATCACCCCGAGTGTGATCGTGCAGCCGCCCTCCGCTGGAAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300

Qy 1012 CCTCTTAACAATCCATGGCATTTGGACGGCATTTGGCGGGATTTTGAGACAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

Qy 1072 TTCTATGAGCGCGGATACCAAGTAGAAAAAGTCAATTTCTCATCCAAATATGATCTCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340

Qy 1132 ACCAAGAACAATGACATTTGGCTGTGATGAGCTGCAGAGCTCTGACTTTCACGACCTTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360

Qy 1192 GTCAAAACCATGTGTCTGCCCAACCCAGCATGTCTGCAGCCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380

Qy 1252 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGAGAACCTCAGAAAGTGTGAAAGCTGGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400

Qy 1312 AAGTGTCTTCTCATTCAGACACAGAGATCCCAACAGCAGATATCTCTATGACACCTGATC 1371
Db 401 LysValLeuLeuLeuGlnThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuLe 420

Qy 1372 ACACCAAGCCATGATCTGTGCGCGCTTCTGCGGGGAAACGTCGATTTTCCCGAGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

Qy 1432 AGTGGAGGCGCTCTGCTCACTTCGAGAACATATCTGCTGGCTGATAGGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460

Qy 1492 TGGGGTCTGCGTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480

Qy 1552 ACGGACTGGATTATTCAGACAAATACAGGGCGACGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

```

RESULT 9

```

US-10-144-678A-932
; Sequence 932, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.

```



```
/ APPLICANT: Jiang, Yuqiu
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Panger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel X.
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skeiky, Yasir A. W.
/ APPLICANT: Hepler, William T.
/ APPLICANT: Hural, John
/ APPLICANT: McNeill, Patricia D.
/ APPLICANT: Houghton, Raymond L.
/ APPLICANT: Vinals y de Bassols, Carlota
/ APPLICANT: Foy, Teresa M.
/ APPLICANT: Watanabe, Yoshihiro
/ APPLICANT: Deng, Ta
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: DIAGNOSIS OF PROSTATE CANCER
/ CURRENT FILING DATE: 2002-08-12
/ NUMBER OF SEQ ID NOS: 1033
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 932
/ LENGTH: 492
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-144-678A-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 14 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-144-678A-932 (1-492)

QY 112 ATGGCTTTGAACCTCAGGGTCACACAGCTATGGACCTTACTATGAAACACCATGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyr-TyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGCCACTGTGTCGCCCACTGCTTAACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCGTCGCCCGTGCAGTACGCCCGGAGGTCCTGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCGTCGTGACGAGCCCAATCCCGATCCGGACAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACTTGAACCTCGGGGACCTTCTCGTGGGAGCTCGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 CGCGCTGGCCTACTCTGGAAGTTCATGGGACGAGTGTCTCCAACTCTGGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyTyrGluCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGGTGTGCACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGAGGACAGAAATCGTGTGTCGCCCTCTACGGACCAACTCATCTCAGGTG 591
|||||
```

RESULT 10

US-10-294-025-932

; Sequence 932, Application US/10294025

```
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAGACTCCTGCGCACCTCTGTGCCAAGACGACTGCAACGAGAACTAC 651
|||||
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GGGCGGGCGGCTGCGAGGACATGGGCTATAGATATATTTTACTCTAGCCAGCAATA 711
|||||
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACCGACTTTATGAACTGAAACACAAAGTCCGCGCAATCTCGAT 771
|||||
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACCAAGTGCCTGCTTCTTCAAAGCGAGTGTCTTCTTACGC 831
|||||
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
QY 832 TGTATAGCTCGGGGCTCAACTTGAACCTCAAGCCGCGCAGACGAGGATGTGGCGCGCAG 891
|||||
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGCGCTGCGCTGCGAGGTCAAGCTGCGACGTCAGAACGTCACGCTG 951
|||||
Db 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCATCACCCCGAGTGGATCGTGACAGCGCCCACTGCGTGAAGAA 1011
|||||
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAAACAATCCATGCGATTGACCGGCATTTCGGGGATTTTGAGACAATCTTTCATG 1071
|||||
Db 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGCGCGGATACCAAGTAGTAAAGTGTCTCTCATCCAAATTTGATCTCCCAAG 1131
|||||
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAAACAATGACATTGCGCTGATGAAGTCGACAGAGCCCTCTGACTTTCAACGACCTA 1191
|||||
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACCAAGTGTCTGCCCAACCGGCGATGATGCTGCGAGCCAGACACACTCTGCTGG 1251
|||||
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCystTrp 380
QY 1252 ATTTCCGGGTGCGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTCTGCAACGCTGCC 1311
|||||
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCATTGACACACAGATGCAACAGCAGATATGCTATGACAACCTGATC 1371
|||||
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCCAGTGTGCGCGGCTCTCTGCGAGGAACTCGATTCTTTCGACGGGTGAC 1431
|||||
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTGGTCACTTCCGAAGACAATATCTGGTGGCTGATAGGGGATCAAGC 1491
|||||
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGGCTGCGCAAGCTTACAGACAGAGTGTACGGAAATGTGATGATGATTC 1551
|||||
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGGATTTATCGACAATAGGGGCGAGACGGC 1587
|||||
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
|||||
```



```

; Publication No. US20030185830A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael B.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210421.427C29
; CURRENT APPLICATION NUMBER: US/10/294, 025
; CURRENT FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-294-025-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 14 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-294-025-932 (1-492)

QY 112 ATGGCTTTGAATCAGGCTCACCACGAGTATTGGACCTTACTATGATAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCAGCCGACCTGTGGTCCGCTCTACGAGGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGCTCCCGCTGCCAGTACGAGGCTCTCTGACGAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCACACCCGCTGCTGTCAGCGAGCCCAATCCCGATCCGGGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCAGCTTACCTCGGAGACCTTCTCGTGGGAGCTGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTACTCTGGAAGTTTCATGGGAGCAAGTGTCTCAACTCTGGGATAGAGTC 471
DB 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACCTGCATCAACCCCTCTAACTGGTGATCGCTGTGCACACTGCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCCGGTGTGTGGCTCTACGGACAAACTTCATCTTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAAGTCTGGCACCCTGTGTGCCAGACGACTGGAACGAGACTAC 651
DB 161 TyrSerSerGlnArgLysSerTyrPheProValCysGlnAspAspTyrAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTTCAGGGACATGGGCTATAAGAAATATTTTACTCTAGCCAAAGGAATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAATCAACACAAAGTCCCGGCAATGTGCAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAGAACTGTACCAGTGTGCTGTTCTTCAAAAGCAGTGGTTCTTTACGC 831

```

RESULT 11

```

US-10-334-038-6
; Sequence 6, Application US/10334038
; Publication No. US20030138865A1
; GENERAL INFORMATION:
; APPLICANT: MacBeth, Kyle J.
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES USEFUL IN
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: 07334-135001
; CURRENT APPLICATION NUMBER: US/10/334,038
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US/09/408,628A
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: US 09/164,159
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: US 09/163,759
; PRIOR FILING DATE: 1998-09-30

```

```

221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGCTCAACTTGAACCTCAAGCCCGCAGAGAGATTCTGGCGCGCAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCTGGCCCTGGCAGGTCAACCTGCAGCTCCAGAACCTCCACGTG 951
DB 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGAGTGGTCTGACAGCCCGCCACTCCCTGGGAAAAA 1011
DB 281 CysGlyGlySerIleIleThrProGluThrPheIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATTCGACCGCATTTGGCGGAGTTTGGCGGAGTTTTCAGACAATCTTTCATG 1071
DB 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCCCGGATACCAAGTACAAAAGAGTATTTCTCATCAAAATATGATCTCCAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTGCGCTGATGAGCTGCGAAGAGCTCTGACCTTTCACACGCTTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACGAGTGTGTCTGCCCAACCCAGGATGATGTGCACGACGACAGCTCTCTCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyrP 380
QY 1252 ATTTCCGGTGGGGGCGCCACCGAGGAGAAAGGAGAGAGCTCAGAAAGTGTCTGAAACGCTGCC 1311
DB 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTCATTGAGACAGAGATGCAACAGACAGATATGTCTATGACAACTGTATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACAGGACCATGATCTGTCCCGCTTCTCGAGGGAGACGTCGATTTCTTGGCAGGCTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTCGTCACTTCGAAAGAACATATCTGTGTGTGTGATAGGGGATACAAAC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
QY 1492 TCGGGTCTGGCTGTGCCAAAGCTTACAGACGAGGAGTGTACGGGAATGTGTGATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGGATTTATCGACAAATGAGGCGACAGCGC 1587
DB 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492

```

NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-334-038-6

Alignment Scores:
Pred. No.: 4,62e-206 Length: 492
Score: 2704.00 Matches: 489
Percent Similarity: 99.80% Conservative: 2
Best Local Similarity: 99.39% Mismatches: 1
Query Match: 85.06% Indels: 0
DB: 14 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-334-038-6 (1-492)

```
QY 112 ATGGCTTGAAGTCAAGGTCACACAGCTATTGAGCACTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGACAGCCCACTGGTCCCACTGTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCCTGCCCGTCCCGACAGTACGCCCGAGGTCCTCAGCAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCACCCCGCTCTCGACGACGACCCCAATCCCATCCGGGACAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACTGACCTTCCCTGGGGACCTTCTCGTGGAGCTGGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTACTCTGGAGTTCATGGGACAGTCTCCAACTCTGGGATGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGlnCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTCTAACTGTGTGTGATGGGTGTCTCACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGAGAGAGATCGGTGTGCTCTACGACCAAACTTCATCCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAGTCTGGGACCTGTGTGCCAAGACACTGGAACAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerThrPheProValCysGlnAspSerThrAsnGluAsnTyr 180
QY 652 GGGCGGGCGCTCGCAGGACATGGGCTATAGAATAATTTTACTCTAGCCAAAGGATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACGCGGATCCAGCTTTATGAAACTGAACACAGTGGCGGCAATGTCAT 771
Db 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAGTGTACACAGTGTGCTGTCTTCAAAAGCAGTGGTCTTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTGGGGGTCACTTGAATCAAGCCGACAGCAGGATGTCGGCGCGAG 891
Db 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCTGGCCCTGGCAGGTGCAGCTGCAGCTCCAGAACCTCCACGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
```

```
QY 952 TGGGAGGCTCCATCATCATCCCGGAGTGGATCGTGACAGCCGCCCACTCGCTGGAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTTTAAACAATCCATGGCATTTGGACGGCATTTGGCGGATTTTGAGACATTTTCATG 1071
Db 301 ProLeuAsnAsnProThrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGAGCCGATACCAAGTAGAAAAAGTGATTCTTCATCCAAATATGACTCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTTGGCTGTGATGAAGCTCCAGAGCTCTGTGACTTTCAAGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTCAAAACCATGTGTCTGCCCAACCCAGGCATCATGTGCAGCCAGAACAGCTCTGTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnCysTyr 380
QY 1252 ATTTCCGGGTGGGGGCCACCCGAGGGAAGAGGAGAACCTCAGAGTGTGACGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCAAGCCATGATGTGTGCCGCTTCTCCAGCGGGAGAACCTCGATTTTCCGAGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCTCTGCTCACTCGAGAACCAATATCTGGTGGCTGTAGAGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleThrPheIleGlyAspThrSer 460
QY 1492 TGGGTTCTCGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGAAATGTGATCGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGATTTATCCAAATGAGGCGCAGACGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
```

RESULT 12

US-09-825-751A-63
; Sequence 63, Application US/09825751A
; Publication No. US20030065140A1
; GENERAL INFORMATION:
; APPLICANT: Curagen Corporation
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Rastelli, Luca
; APPLICANT: Heriman, John L.
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-750
; CURRENT APPLICATION NUMBER: US/09/825,751A
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/194,314
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-825-751A-63

Alignment Scores:
Pred. No.: 8e-206 Length: 492
Score: 2701.00 Matches: 487
Percent Similarity: 99.80% Conservative: 4
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 84.96% Indels: 0
DB: 12 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-825-751A-63 (1-492)

QY 112 ATGGCTTTCACTCAGGTCACACAGCTATTGGACCTTACTATGATAAACCATGGATAC 171
Db 1 MetAlaIeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACGGAAACCCCTATCCGACAGCCCACTGTGTCCTCCACCTGCTTACGAGGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CGGGCTCAGTACTACCGCTCCCGCTGCCCGCCAGTACGCCCGGAGGTCTCGACGAGGT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCGTCGTCGAGGAGCCCAATCCCATCCGACAGTGGACCTCAGAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCAGTGCATCACTTACCTGGGACCTTCTCGTGGGAGGTGGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPhelLeuValGlyAlaLeu 100
QY 412 GCGCTGCGCTACTCTGGAAGTTCATGGGAGCAAGTCTCCACTCTGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGTACTGCATCAACCTCTCACTGCTGATGGCGTGCACACTGCCCT 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCGGTGTGCTTACCGACCAAACTCATCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGGAGTCTGGCCCTCTGTCGACAGACCTGGAACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTyrPheProValCysGlnAspSerPheAsnGluAsnTyr 180
QY 652 GCGCGCGCGCTCGCAGGACATGGGCTATAAGATAATTTTACTCTAGCCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACAGCTTTATGAACCTGAAACAACTGACAGTGGCAATGCGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACCAAGTATGCTGCTTCTTCAAAAGCAGTGGTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATACCTCGGGGTCAACTGAATCAAGCCCGCAGAGCAGGATTTGGCGGGAG 891
Db 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCGGGGCTGGCCCTGCGAGTGCAGCTGACGTCACGACGACGACGTCACG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGCAGGCTCCATCATCACCCCGAGTGGATCTGACGCGCCCACTCGCTGGGAAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACATCATCATGACGCAATTCGGGGGATTTTGACACAATCTTTCATG 1071

RESULT 13
US-09-759-143-895
Sequence 895, Application US/09759143
Patent No. US2002002248A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedwick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C23
CURRENT APPLICATION NUMBER: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 895
LENGTH: 492
TYPE: PRP
ORGANISM: Homo sapiens
US-09-759-143-895

Alignment Scores:

Pred. No.: 2e-205 Length: 492
 Score: 2696.00 Matches: 486
 Percent Similarity: 99.80% Conservative: 5
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 84.81% Indels: 0
 DB: 9 Gaps: 0

US-09-323-597c-1 (1-1738) x US-09-759-143-895 (1-492)

```

QY 112 ATGGCTTTGAAGTACAGGTCACACAGCTATTGACCTTACTATGAAACAGATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaileGlyProTyTyTyGluAsnHisGlyTy 20
QY 172 CAACCGGAAACCCCTATCCCGACAGCCACTGTCGTCCTCCACCTGTCTACAGAGTGAT 231
DB 21 GlnProGluAsnProTyTyProAlaGlnProThrValValProThrValTyGluValHis 40
QY 232 CCGGCTCAGTACTACCGGTCCTCCGTCGCCAGTACGCCCGAGGGTCTTACGAGCGCT 291
DB 41 ProAlaGlnTyTyProSerProValProGlnTyTyAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCGCTGCTGACAGGACGCCAAATCCCATCCGAGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCACTACCTTGACCTTGGGAGCTTCTCTGTCGAGCTCGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGCGCTACTCTGGAGTTCATGGGCACAGTGTCTCAACTCTGGGATAGAGTGC 471
DB 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGAGGACGAGAAATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTATCTCAGAGGAAGTCTGGCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
QY 652 GCGCGGGCGGCTGCGGACATGGGCTATAGAAATAATTTTACTTCTAGCCAAAGATA 711
DB 181 GlyAlaAlaCysArgAspMetGlyTyTyLysAsnAsnPhetyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGGGATCCACACAGCTTTATGAACCTGACACAGTGGCGCAATGTCTGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACTGTACACAGTATGATGCTCTTCTTCAAAAGCAGTGTTTCTTTACGC 831
DB 221 IleTyTyLysLysLeuTyHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCCGCAGAGGAGTGTGGCGCGCGAG 891
DB 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCGCTGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 951
DB 261 SerAlaLeuProGlyAlaTrpProTyProGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATATCCCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CTTCTTAAATTCATGGGATGGAGCGGATTTGGGGATTTGGAGCAATCTTTTCATG 1071
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

```

```

QY 1072 TTCTATGGAGCGGATACCAAGTAGAAAGATGATTTCTTCATCCAAATTTACTCCCAAG 1131
DB 321 PheTyTyGlyAlaGlyTyGlnValGlnLysValIleSerHisProAsnTyAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTTGCGTGTGATGAAGCTGCAGAAAGCTCTGACTTTTCAACGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCCAGTGTGTCTGCCCAACCCAGGATGATGTGTGAGCCAGAACAGCTCTGTCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCCGGTGGGGGCCACCGAGGAAAGGAGGAAAGCTCAGAAAGTGTCTGAAAGCTGCCC 1311
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTTATGAGACACAGAGATGCAACAGCAGATATGTCTATGCAACCTGTATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyValTyAspAsnLeuIle 420
QY 1372 ACACAGCCATGATCTGTGCGGCTTCTCTGACGGGAAACGTCGATTTCTTGCAGGGTAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGAAGGCTCTGTGTCTACTTTCGAAACAATATCTGTGTGTGTGTGTGTGTGTGTGTGT 1491
DB 441 SerGlyGlyProLeuValThrSerAsnAsnAsnIleTrpLeuIleGlyAspHisSer 460
QY 1492 TGGGGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValTyGlyAsnValMetValPhe 480
QY 1552 ACGACTGTGATTTATCGACAAATGAGGCGACAGCGC 1587
DB 481 ThrAspTrpIleTyArgGlnMetLysAlaAsnGly 492

```

RESULT 14

US-09-780-669-895

; Sequence 895, Application US/09780669

; Patent No. US2002005197A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqi

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedrick, Thomas S.

; APPLICANT: Carter, Barrick

; APPLICANT: Wang, AiJun

; APPLICANT: Li, Samuel

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Hepler, William

; APPLICANT: Hural, John

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Houghcon, Raymond L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427C24

; CURRENT APPLICATION NUMBER: US/09/780.669

; CURRENT FILING DATE: 2001-02-09

; NUMBER OF SEQ ID NOS: 943

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 895

; LENGTH: 492

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-780-669-895

Alignment Scores:

Pred. No.: 2e-205 Length: 492
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 84.81% Indels: 0
DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-780-669-895 (1-492)

QY 112 ATGGCTTTGAATCAGGGTCACACAGCTATTGGACCTTACTATCAAAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAACACCCCTATCCCGACAGCCACCTGCTGCTCCCACTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCAACCCCTGCTGTCACGCGACCCCAATCCCGATCCGCGACAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAACACTGTCACTCACCTCACCTCGGGACCTTCTCTGCGGAGCTCGGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTACTCTCGAAAGTTCATGGCGAGCAAGTCTCCAACTCTGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGTACCTCATCAACCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GGGCGGAGGACGAGATCGGTGTGCTGCTCTACGGACCAACTTCTCTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCGAGGAAGTCTGCAACCTGTGTGCGCAAGCACTGGAACGAGACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
QY 652 GGGCGGCGGCTCAGGACATCGGTGTGCTGCTCTACGGACCAACTTCTCTCAGGTG 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
QY 712 GTGGATGACCGGATCCACGCTTTATGAACCTGAACACAACTGCAAGTCCGCGCAATGCT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAACAACGTACCAAGTATGCTGCTCTTCAAGCAGAGTCTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
QY 832 TGATAGCTCGGGGGTCAACTTGAACCTCAAGCCCGCAGACAGGATTTGGCGCGCAG 891
Db 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGCTGCGCTGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTyrProGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCCATCATCACCCCGAGTGGATCGTACAGCCCGCCCACTGCTGCTGCAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGACCGCATTTGGACCGCATTTGGCGGGATTTTGAGACAATCTTTCATG 1071

Db 301 ProLeuAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGGATACCAAGTAGAARAAGTATTTCTCATCCAAATATGACTCCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAATGACATTCGCTGATGAAGCTGCAAGACCTCTCTCACTTTTCAACACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGCTGGGGGCCCAACCGAGGAGAAAGGAGAACTCAGAACTGCTGAACTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTCATGACACAGATGCAACACAGATATGTCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCCATCATCTGTCGGCTTCTGTCAGGGGAACTCGATTCTTCCAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGAAGGCTCTGCTCACTTCGAAGAACAAATATCTGCTGCTGATAGGGGATACAAG 1491
Db 441 SerGlyGlyProLeuValThrSerAsnAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGCTGATTTATCGACAAATGAGGCGAGACGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetLysAlaAsnGly 492

RESULT 15

US-09-879-792-14
Sequence 14, Application US/09879792
Patent No. US20020061850A1
GENERAL INFORMATION:
APPLICANT: Xiao, Yonghong
APPLICANT: Gedrich, Richard
TITLE OF INVENTION: Regulation of Human Transmembrane Serine
TITLE OF INVENTION: Protease
FILE REFERENCE: 02973.00035
CURRENT APPLICATION NUMBER: US/09/879,792
CURRENT FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/211,224
PRIOR FILING DATE: 2000-06-13
PRIOR APPLICATION NUMBER: US 60/283,353
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: US 60/283,648
PRIOR FILING DATE: 2001-04-16
PRIOR APPLICATION NUMBER: PCT
PRIOR FILING DATE: 2001-06-12
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-879-792-14

Alignment Scores:
Pred. No.: 2e-205 Length: 492
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 84.81% Indels: 0

{Docket No. US20020061850A1 LIO-81-WO}

DB: 9 Gaps: 0
US-09-323-597C-1 (1-1738) x US-09-879-792-14 (1-492)
QY 112 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGATAAACCAGTGATAC 171
Db 1 MetAlaLeuAenSerGlySerProProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGTCCCACTGTCTACGAGTGCAT 231
Db 21 GlnProGluAenProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCGTCGCCCGTGCACAGTACGCCCGAGCGTCTCTGACGACGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTGTGTCACGAGCCCAAAATCCCATCCGGGACAGTGTGTCACCTCAAAG 351
Db 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCTCGCGGACCTTCTCGTGGGAGCTGGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCCCTGCGCTACTCTGGAATTCATGGGACGCAAGTCTCCAACTCTGGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAenSerGlyIleGluCys 120
QY 472 GACTCCTCAGCTACTGCATCAACCCCTCACTGCTGTGATGGCTGCACACTGCCCC 531
Db 121 AspSerGlyThrCysIleAenProSerAenThrCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGAGGACGAGATCGGTGTGTGCTCTACGGACCAAACTTCATCTTCAGGTG 591
Db 141 GlyGlyGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGGAGTCTTGGACCCCTGTGTGCGGACGACGACGACGACGACGAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAenAspTrpAenGluAenTyr 180
QY 652 GCGCGGCGCGCTGCAGGACATGGCTATAGAAATAATTTTACTTAGCCAAAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACACGGGATCCACAGCTTTATGAACTGAACACAAAGTGGCGGCAATGTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp 220
QY 772 ATCTATAAAACCTGACACAGTGTGCTCTCAAGCCGACAGTGGTTCCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCAACTTGAATCAAGCCGACGACGACGACGACGACGACGACGAC 891
Db 241 CysLeuAlaCysGlyValAenAenAenSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCGGGGCTGGCCCTGCGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 951
Db 261 SerAlaLeuProGlyAlaIleProTyrProGlnValSerLeuHisValGlnAenValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGGAGTGTGTCAGCAGCGCCGCTGCTGCTGCTGCTGCT 1011
Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACATCATGCGCATGACGCGCATTTGGGGGATTTGAGACATCTTTCATG 1071
Db 301 ProLeuAenAenProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCATGAGCGCGGATACCAAGTAGAAAGTGTATTTCTATCCAAATTTATGACTCCAAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAenTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTTGGCTGTGATGAAGCTGACGAGGCTCTGACTTTCACGACCTA 1191

Search completed: July 6, 2004, 15:23:10
Job time : 117 secs

Db 341 ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAenAspLeu 360
QY 1192 GTCAAAACCAAGTGTGTGCCCCAACCCAGGATGATCTGTGACGCCAGAACAGCTCTGTCTGG 1251
Db 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCCCAGGAGGAGAAAGGAGAACCTCAGAAAGTGTGAACGCTGCC 1311
Db 381 IleSerGlyTyrPheAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAla 400
QY 1312 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATCTATGACAAACCTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAenAenLeuIle 420
QY 1372 ACACCGACCATGATCTGTGCGGCTTCTCCAGGGGAAAGTGTGATTTCTTCCAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGGCTCTGTGCTCACTTCGAAAGAAACAATATCTGTGGTGTGATAGGGGATACAAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerAenAenAenIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGTGCTGCCAAAGCTTACAGACGAGGAGTGTACGGGAATGTGATGCTATTTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480
QY 1552 ACGGACTGGATTTTATCGACAAATGAGGCGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetLysAlaAenGly 492

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 14:59:51 ; Search time 24 Seconds
(without alignments)
7477.161 Million cell updates/sec

Title: US-09-323-597C-1

Perfect score: 3179
Sequence: 1 ggcggagggcgagcgagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p-model -DEV=xlp
-Q/cgn2_1/USPTO.spool_P/US09323597/runat_06072004_111728_29711/app_query.fasta_1.1927
-DB=Issued Patents AA -CPMT=fastan -SUPPLX=rai -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09323597 -CGN1_1_27 -runat_06072004_111728_29711 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCKS=100 -LONGLOG
-DE TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2717	85.5	492	3	US-09-342-749-2
2	2717	85.5	492	4	US-09-691-840-2
3	2696	84.8	492	4	US-09-685-166A-895
4	1540	48.4	283	3	US-08-807-151-1
5	1540	48.4	283	4	US-09-478-957-1
6	1165	36.6	209	4	US-09-685-166A-897
7	888	27.9	454	3	US-09-518-046-2
8	866	27.2	159	3	US-09-518-046-24
9	780.5	24.6	455	3	US-09-261-416-2
10	684	21.5	423	4	US-09-656-002-2
11	676.5	21.3	406	4	US-09-851-588-6
12	676.5	21.3	435	3	US-09-008-271A-6

13	676.5	21.3	437	4	US-09-851-588-8	Sequence 8, Appli
14	660	20.8	798	1	US-08-200-900A-2	Sequence 2, Appli
15	660	20.8	798	5	PCT-US94-00616-2	Sequence 2, Appli
16	575.5	18.1	855	2	US-09-027-337-2	Sequence 2, Appli
17	575.5	18.1	855	4	US-09-644-600-2	Sequence 2, Appli
18	575.5	18.1	855	4	US-09-654-600A-2	Sequence 2, Appli
19	567	17.8	417	4	US-09-820-002-4	Sequence 4, Appli
20	558.5	17.6	376	4	US-09-820-003-2	Sequence 2, Appli
21	556	17.5	638	2	US-08-681-151-3	Sequence 3, Appli
22	553.5	17.4	248	3	US-08-944-483-63	Sequence 63, Appli
23	553.5	17.4	812	1	US-08-248-629A-1	Sequence 1, Appli
24	553.5	17.4	812	1	US-08-451-932-1	Sequence 1, Appli
25	553.5	17.4	812	1	US-08-452-260-1	Sequence 1, Appli
26	553.5	17.4	812	1	US-08-326-785-1	Sequence 1, Appli
27	553.5	17.4	812	2	US-08-612-788-1	Sequence 1, Appli
28	553.5	17.4	812	2	US-08-605-598B-1	Sequence 1, Appli
29	553.5	17.4	812	2	US-08-429-743-1	Sequence 1, Appli
30	553.5	17.4	812	2	US-08-866-735-1	Sequence 1, Appli
31	553.5	17.4	812	3	US-09-066-028-1	Sequence 1, Appli
32	553.5	17.4	812	4	US-09-192-012-3	Sequence 3, Appli
33	553.5	17.4	812	4	US-09-335-325-1	Sequence 1, Appli
34	553.5	17.4	812	4	US-08-991-761A-12	Sequence 12, Appli
35	553.5	17.4	812	5	PCT-US95-05107-1	Sequence 1, Appli
36	548.5	17.3	902	4	US-09-644-600-10	Sequence 10, Appli
37	548.5	17.3	902	4	US-09-654-600A-10	Sequence 10, Appli
38	541	17.0	416	2	US-09-000-845-2	Sequence 2, Appli
39	534	16.8	356	2	US-08-681-151-1	Sequence 1, Appli
40	528.5	16.6	256	2	US-09-027-337-3	Sequence 3, Appli
41	528.5	16.6	256	4	US-09-644-600-3	Sequence 3, Appli
42	528.5	16.6	256	4	US-09-654-600A-3	Sequence 3, Appli
43	527.5	16.6	418	1	US-08-508-448C-25	Sequence 25, Appli
44	527.5	16.6	418	1	US-09-370-838-62	Sequence 62, Appli
45	527.5	16.6	418	4	US-09-370-838-82	Sequence 82, Appli

ALIGNMENTS

RESULT 1
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-P.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Alignment Scores:			
Pred. No.:	2,376-232	Length:	492
Score:	2717.00	Matches:	492
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	85.47%	Indels:	0
DB:	3	Gaps:	0

US-09-323-597C-1 (1-1738) x US-09-342-749-2 (1-492)

QY 112 ATGGGCTTTCAACTCAGGGTCCACCAGCTATTGCACCTTACTATGAACCATGATAC 171
|||||
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
|||||


```
QY 172 CAACGGGAAACCCCTATCCCGGACAGCCCACTGGTCCCTCACTGCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CGGCTCAGTACTACCGGTCCTCCCGTGCCTCCAGTACCGCCCGAGGTCCTCAGCAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTCAGCAGCAGCCCAATCCCATCCGGGAGCAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGACTGTGCATCACTTGACCTCGGACCTTCCTCGGAGAGCTGCGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 CGCGTGCCTACTCTGGAAGTTCATGGGACAGAGTCTCCAACTCTGGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGTACTGCATCAACCCCTCACTGCTGATGCTGATGCTGATGCTGATGCTG 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCGGTGTGCTGCTGCTACGACCAAACTTCATCTTCAGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAACTCTGACACCTCTGTCGCAAGACGACTGGAACGAGACTAC 651
Db 161 TyrSerSerGlnAsGlySerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGGGGGGCTCAGGACATGGCTATAGATATATTTTACTCTAGCCAGGAGATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
QY 712 GTGATGATCAGCGGATCCACAGCTTTATGAACCTGAACACAAAGTCGCGCAATGTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAATACTGTACACAGTATGCTGCTCTTCAAAAGCAGTGGTCTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGCTCAACTTCAACTCAAGCGCGCAGACGAGATGTGGCGGCGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCGGGGGCTGCGCTGCGCAGCTGAGCTGACGTCGACGACGTCACGCTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCACCGAGTGGATCGTACAGCGCGCCACTGCTGCGAAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACATCCATGCACTGACGCGCATTTGCGGGATTTTGAGACAATCTTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrPheTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGGATCCCAAGTAGAAAAGTATTCTCTCCAAATTTAGCTCCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAAACAATGACATTCGCTGATGAGCTGACAGAGCTCTGACTTTTCAAGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACAGTGTCTGCTCCCAACCCAGGATGATGCTGACGACGACAGACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
```

```
QY 1252 ATTTCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTTCAGAGTGTGAAGCTGGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTGAGACACAGATGCAACAGCAGATATCTCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGACCATGATCTGTGCGGCTTCTGCGAGGAAAGTCGATTTCTTCCGAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCTCTGCTGCTCAAGAACAAATATCTGTGGCTGATAGCGGATACAAGC 1491
Db 441 SerGlyCysProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGGCTGCCAAAGCTTACAGACCGAGGTGTACGGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGGATTTATCGACAAATGCGGCGCAGCGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 644419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-P.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-691-840-2

Alignment Scores:
Pred. No.: 2,378-232 Length: 492
Score: 2717.00 Matches: 492
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 85.47% Indels: 0
DB: 4 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-691-840-2 (1-492)
QY 112 ATGGCTTTGAACCTCAGGCTCACCACGAGCTATTGGACCTTACTATGAAAACCATGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGGACAGCCCACTGGTCCCTCACTGCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CGGCTCAGTACTACCGGTCCTCCCGTGCCTCCAGTACCGCCCGAGGTCCTCAGCAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTCAGCAGCAGCCCAATCCCATCCGGGAGCAGTGTGCACCTCAAG 351
```

Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCAGTGTGCTACCTTGACCTGGGAGCACTTCTCTGGGAGCTGGCTG 411
Db 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCGTGGCTACTCTGGAGTTTCAGGGCAGCAAGTGTCTCAACTCTGGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTTCAGTACCTGCACTCAACCCCTCTAACTGTGTGATGGTGTGTCACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGAGACAGCAATCGTGTCTGCTCTAGGACCAACTTCATCTTCAGGTG 591
Db 141 GlyGlyLeuAspGluAsnArgCysValArgLeuTrpGlyProAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAGTCTCGCACCTCTGTGCGCAAGACGACTGGAACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
QY 652 GCGCGGCGGCTCGCAGGACATGGGCTATAGCAATATTTTACTCTAGCCAGCAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGGATCCACGACTTATGAACTGAACACAGTGGCGGCAATGTCGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAATACTGACACAGTGTGCTGCTCTTCAAAAGCAGTGGTCTTTTACCC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGNATAGCTCGCGGGTCAACTGCACTCAAGCGCGCAGACGAGTGTGGCGCGCAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgLysSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGCGCTGGCCTGGCAGGTGAGCTGACCTGCCAGAACGTCACGCTG 951
Db 261 SerAlaLeuProGlyAlaTrpTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCCATCATCCCCCGAGTGATGTGACAGCGCCGCTCGTGGGAAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACCAATCCATGGACGTCATTTGGCGGCAATTTGGACAACTTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGACCGGATACCAAGTAGAAGAAAGTGTCTTCATCCAAATGACTCCAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTrpAspSerLys 340
QY 1132 ACCAAGAACCAATGACATTCGCTGATCAAGCTGACAGAGCTCTGACTTTTCAACGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACAGTGTCTGCCCAACCCAGGCAATGATGTGAGCAGAACGACTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCACCGAGGAAGGAGACCTCAGAAGTGTGGAAGCTGCC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTCATGTAGACACAGAGATGCAACAGCAGATATGCTATGACACCTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCGACGATGCTGTGGGCTTCTCGCAGGGAACCTTCGATCTTCCAGGCTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

QY 1432 ACTGAGGCGCTCGTCTCACTTCGAGAACCAATATCTGCTGCTGATAGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTTCTGCTGTGCCAAAGCTTACAGACGAGAGTGTACGGGAATGTGATGTTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTCGATTATCGACAAATGAGGCGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 3

US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, John A.
; APPLICANT: Stolk, John W.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Alignment Scores:
Pred. No.: 1,72e-230 Length: 492
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 84.81% Indels: 0
DB: Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-685-166A-895 (1-492)

QY 112 ATGGCTTCAACTCAGGCTCACCACGACTATTGGACCTTACTATGAAAACCATGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTrpGluAsnHisGlyTyr 20
QY 172 CAACCGAAAAACCCCTATCCCGCAGACCCCACTGTGGTCCCACTCTCTACGAGGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CGGGCTCAGTACTACCGTCCCGCTCCCGCTAGTACGCCCGCGGCTCTGTGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCGCTCGTGTGACCGCAGCCCAATCCCATCCGCGACAGTGTGCACTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

QY 352 ACTAAGAAAGCAGCTGTGCATCACCTTGACCCCTGGGAGACCTTCCCTCGTGGAGCTGCCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCCCTGCGCTACTCTGGAAGTTCATGGGAGCAAGTCTCCAACTCTGGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGTACTCTGCATCAACCCCTCACTTAACTGGTGTGTGATGGGTGACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAspTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCGGTGTGCTCGCTCTACGACCAAACTTCATCTCCAGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGAACTCTGGCACCCCTGTGTGCCAAGACGACTGGAAGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGGGGCGGCTCGCAGGACATGGCTTATGAATATTTTACTCTAGCCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACGAGCTTTATGAACCTGAACAACTGCGGCAATGTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAGTGTACACAGTATGCTGCTTCTTCAAAAGCAGTGGTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGCTCAACTTGAACCTCAAGCGCCGACAGCAGGATTTGGCGGGGAG 891
Db 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCGCGGGGCTCGCCCTCGCAGCTGAGCTGACGTCAGCAGCAGCAGCTCAGTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGCGAGGCTCATCATCACCCCGAGTGGATCGTGACAGCGCGCCACTGCTGGAATAA 1011
Db 281 CysGlyGlySerIleIleThrProGlnTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGCGATGGACGCGCATTTGGGGGATTTTGAGCAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGGATCCAACTAGAAAAGTGTCTCTCATCCAAATTATGACTCCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAAACATGATTCGCTGATGAGTGCAGAGCCCTCTGACTTTCACAGCACTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACAGTGTCTGCCCAACCCAGCATGATGCTGCGAGCCAGACACCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGGCCACCGAGGAGAAAGAGACCTCAGAACTGCTGAGCGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTCATTTGAGACACAGAGATGCAACAGCATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCATGATCTGTGGCGGCTTCTGACAGGGAACGCTGATTTCTGACAGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

QY 1432 AGTGGAGGCGCTTGCTCACTTCGAAGAACATATCTGCTGGCTGATAGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerAsnAsnAsnIleTrpIleuIleGlyAspThrSer 460
QY 1492 TGGGGTTCGTGGCTGCTCCAAAGCTTACAGACCAAGGCTTACGGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACCGACTGATTTATCGACAAATGAGGCGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetLysAlaAsnGly 492

RESULT 4
US-08-807-151-1
; Sequence 1, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO701
; CLONE: 556016
; US-08-807-151-1

Alignment Scores:
Pred. No.: 3,71e-128 Length: 283
Score: 1540.00 Matches: 282
Percent Similarity: 99.65% Conservative: 0
Best Local Similarity: 99.65% Mismatches: 1
Query Match: 48.44% Indels: 0
DB: 3 Gaps: 0

US-09-323-597C-1 (1-1738) x US-08-807-151-1 (1-283)

QY 739 ATGAACCTGAACACAACTGCGCAATGTCTATATCTATATAAACTGTACACAGTGTAT 798
Db 1 MetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLysLeuTyrHisSerAsp 20

```
QY 799 GCGTGTCTTCAAAAGCAGTGGTTCTTTACGCTGTATAGCCTGGGGCTCAACTTGAAC 858
Db 21 AlaCysSerSerLysAlaValValSerLeuArgCysLeuValAsnLeuAsn 40
QY 859 TCAAGCCGCGAGAGAGATTTGGCGCGGAGAGCGGCTCCCGGGCGCTGGCCCTGG 918
Db 41 SerSerArgGlnSerArgLeuValGlyGlyGluSerAlaLeuProGlyAlaTrpProTrp 60
QY 919 CAGGTGAGCTGTCAGCTCCAGAACGTCACGCTGGGGAGGCTCCATCATCACCCCGAG 978
Db 61 GlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIleIleThrProGlu 80
QY 979 TGGATCGTGCAGCCGCCCTCGCTGGGAAACCTCTTAAACATTCATGCGATTGGAGC 1038
Db 81 TrpIleValThrAlaAlaHisCysValGluLysProLeuAsnProTrpHisTrpThr 100
QY 1039 GCATTTGCGGGGATTTGAGCAATCTTTCATGCTCTATGGAGCGGATACCAAGTAGAA 1098
Db 101 AlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGlu 120
QY 1099 AAAGTGATTTCTCATCCAAATATTGACTCCAGACCAAGAACAAATGACATTCGCTGATG 1158
Db 121 LysValIleSerHisProAsnTyrAspSerLysThrLysAsnAspIleAlaLeuMet 140
QY 1159 AAGCTGCAGAGAGCTCTGACTTTCACGACCTAGTGAACCAAGTGTGTGCCCAACCA 1218
Db 141 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 160
QY 1219 GGCATGATGCTGCAGCCAGAACAGCTCTGCTGATTTCCGGGTGGGGCCGCCAGGAG 1278
Db 161 GlyMetLeuGlnProGluGlnLeuCysTyrIleSerGlyTyrGlyAlaThrGluGlu 180
QY 1279 AAGGGAAGACCTCAGAGTGTGAAAGCTGCGAGGTCTCTCATTTGAGACACAGAA 1338
Db 181 LysGlyLysThrSerGluValLeuAsnAlaLysValLeuIleGluThrGlnArg 200
QY 1339 TGCACACAGCAGATATCTATGACAACTGATCACCAGCCATGATCTGTGCGGCTTC 1398
Db 201 CysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGlyPhe 220
QY 1399 CTGACGGGAGCTCGATTCTTCCAGGTGACAGTGGAGGGCTCTGTCTACTTCGAGAG 1458
Db 221 LeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGly**LeuValThrSerLys 240
QY 1459 AACAAATATCTGCTGTGATAGGGATACAGCTGGGTTCTGCTGTGCCAAGCTTAC 1518
Db 241 AsnAsnIleTrpIleLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAlaTyr 260
QY 1519 AGACACGAGTGTACGGGAATGTGATGATTACAGCACTGGATTTATCGACAAATGAGG 1578
Db 261 ArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMetArg 280
QY 1579 GCAGACGCC 1587
Db 281 AlaAspGly 283
```

RESULT 5

US-09-478-957-1

Sequence 1, Application US/09478957

Patent No. 635048

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Lal, Preeti

TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSER: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/478,957
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,151
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4186
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNO01
CLONE: 556016
US-09-478-957-1
```

Alignment Scores:

Pred. No.:	3,71e-128	Length:	283
Score:	1540.00	Matches:	282
Percent Similarity:	99.85%	Conservative:	0
Best Local Similarity:	99.85%	Mismatches:	1
Query Match:	48.44%	Indels:	0
DB:	4	Gaps:	0

US-09-323-597C-1 (1-1738) x US-09-478-957-1 (1-283)

```
QY 739 ATGAACCTGAACACAGTCCGCGCAATGTCGATATCTATAAAACTCTACACAGTGTAT 798
Db 1 MetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLysLeuTyrHisSerAsp 20
QY 799 GCCTGTCTCTCAAAGACAGTGGTTCTTTACGCTGTATAGCTGCGGGCTCAACTTGAAC 858
Db 21 AlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGlyValAsnLeuAsn 40
QY 859 TCAAGCCGCGAGAGAGATTTGGCGCGGAGAGCGGCTCCCGGGCGCTGGCCCTGG 918
Db 41 SerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGlyAlaTrpProTrp 60
QY 919 CAGGTGAGCTGTCAGCTCCAGAACGTCACGCTGGGAGGCTCCATCATCACCCCGAG 978
Db 61 GlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIleIleThrProGlu 80
QY 979 TGGATCGTGCAGCCGCCCTCGCTGGGAAACCTCTTAAACATTCATGCGATTGGAGC 1038
Db 81 TrpIleValThrAlaAlaHisCysValGluLysProLeuAsnProTrpHisTrpThr 100
QY 1039 GCATTTGCGGGGATTTGAGCAATCTTTCATGCTCTATGGAGCGGATACCAAGTAGAA 1098
Db 101 AlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGlu 120
QY 1099 AAAGTGATTTCTCATCCAAATATTGACTCCAGACCAAGAACAAATGACATTCGCTGATG 1158
Db 121 LysValIleSerHisProAsnTyrAspSerLysThrLysAsnAspIleAlaLeuMet 140
QY 1159 AAGCTGCAGAGAGCTCTGACTTTCACGACCTAGTGAACCAAGTGTGTGCCCAACCA 1218
Db 141 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 160
QY 1219 GGCATGATGCTGCAGCCAGAACAGCTCTGCTGATTTCCGGGTGGGGCCGCCAGGAG 1278
```

Db 161 GlyMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTrpGlyAlaThrGlu 180
QY 1279 AAAGGGAGAGCTCAGAGTGTGAACCTGCGCAGGCTTCTCATTGACACACAGAGA 1338
Db 181 LysGlyLysThrSerGluValLeuAsnAlaAlaValLeuLeuIleGluThrGlnArg 200
QY 1339 TGCAACAGCAGATATGTTATGACAACTGATCACCAGCCATGANTGTGCGCGCTTC 1398
Db 201 CysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGlyPhe 220
QY 1399 CTGACGGGAGAGCTGAGTCTTTCAGAGGTGACAGTGCAGGCGCTCTGCTCACTTCGAG 1458
Db 221 LeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGly**LeuValThrSerLys 240
QY 1459 AACATATCTGTGCTGATAGGGATACAAAGCTGGGTTCTGGCTGTGCCAAAGCTTAC 1518
Db 241 AsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAlaTyr 260
QY 1519 AGACAGGAGTGTACGGGAATGTGATGTATTCACGGAGCTGGATTATCGCAATGAGG 1578
Db 261 ArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMetArg 280
QY 1579 GCAGACGGC 1587
Db 281 AlaAspGly 283

RESULT 6

US-09-685-166A-897
; Sequence 897, Application US/09685166A
; Patent No. 6630305

GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqi

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Panger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedwick, Thomas S.

; APPLICANT: Carter, Derrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skelky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C21

; CURRENT APPLICATION NUMBER: US/09/685,166A

; CURRENT FILING DATE: 2000-10-10

; NUMBER OF SEQ ID NOS: 898

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 897

; LENGTH: 209

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-685-166A-897

Alignment Scores:

Pred. No.: 5,58e-95 Length: 209
Score: 1165.00 Matches: 208
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.52% Mismatches: 0
Query Match: 36.65% Indels: 0
DB: 4 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-685-166A-897 (1-209)

QY 112 ATGGCTTTGAACTCAGGCTCACCACGAGCTATTGGACCTTACTATCAAAACCATGTATC 171

Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAAAACCCCTATCCCGCACAGCCACACTGTGGTCCCGCTCTCTACGAGGTCCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CGGCTCAGTACTACCGTCCCGGTCGCCCGCCAGTACGCCCGGAGGTCTCTGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTGACGAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAACAGCTGTGCATCCTTGCACCTCGGGGACCTTCTCTGTTGGGAGCTGCCCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTTACTCTGGAAGTTTCATGGGACAGAGTCTCAACTCTGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCTCAGGTACTCTGCATCAACCCCTCTAACTGTGTGATGGCGTGTGCACCTGCCCT 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGAGGACGAGGANTCGTGTTCGCTCTACGGACCAACTCTCTGATCCTTCAAGTGC 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGGAAGTCTGGCACCTCTGTGTGCCAAGACGACTGGAAAGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTGCAGGACATGGGCTATAGGATATTTTACTTCTAGCCAGGAGATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPhenTyrSerSerGlnGlyLe 200
QY 712 GTGGATGACAGCGGATCCACCGCTTT 738
Db 201 ValAspAspSerGlySerThrSerPhe 209

RESULT 7

US-09-518-046-2

; Sequence 2, Application US/09518046

; Patent No. 6294663

GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed

; FILE REFERENCE: D6192CIP

; CURRENT APPLICATION NUMBER: US/09/518,046

; CURRENT FILING DATE: 2000-03-02

; EARLIER APPLICATION NUMBER: 09/261,415

; EARLIER FILING DATE: 1999-03-03

; NUMBER OF SEQ ID NOS: 153

; SEQ ID NO 2

; LENGTH: 454

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: complete amino acid sequence of TADG-12

US-09-518-046-2

Alignment Scores:

Pred. No.: 2,57e-70 Length: 454
Score: 888.00 Matches: 190
Percent Similarity: 59.02% Conservative: 52
Best Local Similarity: 46.34% Mismatches: 148
Query Match: 27.93% Indels: 20

```

DB:          3          Gaps:          10
US-09-323-597C-1 (1-1738) x US-09-518-046-2 (1-454)
QY 379 ACCTGGGGACCTTCCTCGTGGAGCTGCGCTGGCGCTGCTACTCTGGAAGTTTCATG 438
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 54 SerLeuGlyllelleAlaLeuileLeuAlaLeuAlaAlaGlyLeuGlylleHisPhe--- 72
QY 439 GCGAGCAAGTGTCTCAACTCTGGATAGAGTGGACTCTCTCAGGTACCTGCATCAACCCC 498
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 73 -----AspCysSerGlyLys---TyrArgCysArgSerSerPheLysCysilleGluLeu 89
QY 499 TCTAAGTGTGTGATGCTGCTACACTGCCCGGGGAGGAGACGAGATCGTGTGTT 558
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 90 IleThrArgCysAspGlyValSerAspCysLysAspGlyGluAspGluTyrArgCysVal 109
QY 559 CGCTCTACGACCAAACTTCATCTTCAGGTGACTCATCTCAGAGAAAGTCTCTGCAC 618
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 110 ArgValGlyGlyGlnAsnAlaValLeuGlnValPheThrAla-----AlaSerTrpLys 127
QY 619 CCTGTGTCCAGACGACTGGAACGAGACTAGCGGGCGGCGCTGCAGGGACATGGGC 678
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 128 ThrMetCysSerAspGlyTrpLysGlyHisTyrAlaAsnValAlaCysAlaGlnLeuGly 147
QY 679 TATAAGAAATATTTTCTCTAGCAAGGAATAGTGATGACGAGCGGATCCACCGCTTT 738
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 148 PhePro---SerTyrValSerSerAspAsnLeuArgValSerSerLeuGluGlyGlnPhe 166
QY 739 ATGAACCTGAACACAGTGGCGCAATGTC-----GATATCTATAAAAGTCTAC 789
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 ArgGluGluPheValSerIleAspHisLeuLeuProAspAspLysValThrAlaLeuHis 186
QY 790 CACAGT-----GATGCTCTTCTTCAAAAGCAGTGGTCTTCTTACGCTGATA 837
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 HisSerValTyrValArgGluGlyCysAlaSerGlyHisValThrLeuGlnCysThr 206
QY 838 GCCTGCGGGGTCAACTGAACTCAAGCGCCAGAGCAGATGTGGCGGCGAGAGCGCG 897
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 207 AlaCysGlyHisArgGlyTyrSer-----SerArgIleValGlyGlyAsnMetSer 224
QY 898 CTCGCGGGGCTGGCGCTGACGTGACCTGACGCTCCAGACGCTCCAGCTGCGGA 957
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 LeuLeuSerGlnTrpTrpGlnAlaSerLeuGlnPheGlnGlyTyrHisLeuCysGly 244
QY 958 GGTCTCCATCAPACCCCGAGTGGATCGTGACGCGCCCACTCGCTGGGAAAAACCTCTT 1017
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 GlySerValIleThrProLeuTrpIleThrAlaHisCysVal---TyrAspLeu 263
QY 1018 AACATCCATGGCATTTGACGCGCATTTGCGGGATTTTGCACAAATCTTTCATGTTCTAT 1077
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 TyrLeuProLysSerTrpThrIleGlnValGlyLeuVal-----SerLeuLeuAspAsn 281
QY 1078 GGAGCC---GGATACCAAGTAGAAAAGTGAATTTCTATCCAAATTTATGACTCCAGACC 1134
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 282 ProAlaProSerHisLeuValGluLysIleValTyrHisSerLysTyrLysProLysArg 301
QY 1135 AAGAACATGATGCTGCTGATGAGCTGACGAGAGCTCTGACTTTTCAACGACTAGTG 1194
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 302 LeuGlyAsnAspIleAlaLeuMetLysLeuAlaGlyProLeuThrPheAsnGluMetIle 321
QY 1195 AAACCAAGTGTCTGCCCAACCCAGGCATGATGTGCGAGCCAGAACACTCTGCTGGATT 1254
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 322 GlnProValCysLeuProAsnSerGluGluAsnPheProAspGlyLysValCysTrpThr 341
QY 1255 TCCGGGTGGGGGCCCCAGAGAGAAAGGAGACCTCAGAAAGTGTGAAGCTGCCAAG 1314
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 342 SerGlyTrpGlyAlaThrGluAspGlyGlyAspAlaSerProValLeuAsnHisAlaAla 361
QY 1315 GTGCTTCTCATTTGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATCACA 1374
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 362 ValProLeuIleSerAsnLysIleCysAsnHisArgAspValTyrGlyLysIleIleSer 381
QY 1375 CCAGCCATGATCTGTGCGCGCTTCTGCGAGGGGACGCTGATTTCTTCCGCGGTGACAGT 1434

```


QY 1336 AGATGCAACAGCAGATATGTTCTATGACAACTGATGATCAGCAGCCAGCATGATCTGTGCGGC 1395
Db 121 ArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGly 140
QY 1396 TTCTGCGAGGAGAGCTGATCTTCCAGGGTGCAGTGCAGGGCCCTCTGCTCACT 1452
Db 141 PheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyProLeuValThr 159

RESULT 9
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TAGD-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2

Alignment Scores:
Pred No.: 8, 62e-61 Length: 455
Score: 780.50 Matches: 180
Percent Similarity: 55.6% Conservative: 50
Best Local Similarity: 43.5% Mismatches: 158
Query Match: 24.5% Indels: 25
DB: 3 Gaps: 13

US-09-323-597C-1 (1-1738) x US-09-261-416-2 (1-455)

QY 379 ACCCTGGGAGCTTCTCTGCGGAGCTGGCTGGCGGCTGCTACTCTGGAAGTTCTATG 438
Db 54 SerLeuGlyIleAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuGlyLeuGlyIleHisPhe--- 72
QY 439 GCGAGCAAGTCTCCAACTCTGGGATAGAGTGGCACTCTCAGCTACTGCTGATCAACCCC 498
Db 73 -----AspCysSerGlyLys---TyrArgCysArgSerPheLysCysIleGluLeu 89
QY 499 TCTAACTGGTGTGATGGCTGTACACTGCCCCGGGGAGGACGAGAAATCGGTGTGT 558
Db 90 IleThrArgCysAspGlyValSerAspCysLysAspGlyGluAspGluTyrArgCysVal 109
QY 559 CGCTCTACGACCAAACTTCTATCTTCAGGTGTACTCTATCTCAGAGGAAGTCTCTGGCAC 618
Db 110 ArgValGlyGlyGlnAsnAlaValLeuGlnValPheThrAla-----AlaSerTrpLys 127
QY 619 CCTGTGTCAGGACGATGAAACGAGAACTACGGCGGGCGGCTGCGAGGACATGGCC 678
Db 128 ThrMetCysSerAspTrpLysGlyHisTyrAlaAsnValAlaCysAlaGlnLeuGly 147
QY 679 TATAAGATAATTTTACTCTAGCAAGAAATAGTGGATGACAGGATCCACCACTTT 738
Db 148 PhePro-----SerTyrValSerAspAsnLeuArgValSerSerLeuGluGlyGlnPhe 166
QY 739 ATGAACCTGAACACAAAGTGGCGCAATGTC-----GATATCTATAAAAAAAGTGTAC 789
Db 167 ArgGluGluPheValSerIleAsnHisLeuLeuProAspAspLysValThrAlaLeuHis 186
QY 790 CACAGT-----GATGCTGTCTTCAAAAGCAGTGGTTCTTTACGCTGTATA 837
Db 187 HisSerValTyrValArgGluGlyCysAlaSerGlyHisValValThrLeuGlnCysThr 206

QY 838 GCTTGGGGGTCAACTTGAATCAAGCCCGCAGAGCAGGATTTGGGCGCGCAGAGCGCG 897
Db 207 AlaCysGlyHisArgArgGlyTyrSer-----SerArgIleValGlyGlyAsnMetSer 224
QY 898 CTCGCGGGGCTGGCCCTGGCAGGTGAGCTCCAGCTCCAGAACGTCACAGGTGTCGCGA 957
Db 225 LeuLeuSerGlnTrpProTrpGlnAlaSerLeuGlnPheGlnGlyTyrHisLeuCysGly 244
QY 958 GGTCTCATCATCAACCCCGAGTGGATCTGACAGCGCCGCTGCTGCGGAAAAACCTCTT 1017
Db 245 GlySerValIleThrProLeuTrpIleThrAlaAlaHisCysVal---TyrAspLeu 263
QY 1018 AACAAATCCATGCGCAATGGAGCGCATTTGGGGGATTTTGAGACAATCTTTCATGTTCTAT 1077
Db 264 TyrLeuProLysSerTrpTrpIleGlnValGlyLeuVal-----SerLeuLeuAspAsn 281
QY 1078 GAGAGC---GGATACCAAGTAGAAAAGTATTTCTCATCAAAATATGATCTCCAGACC 1134
Db 282 ProAlaProSerHisLeuValGluLysIleValTyrHisSerLysTyrLysProLysArg 301
QY 1135 AAGAAACAATGACATTCGGCTGATGAAGCTGCAGAGCCTCTGACTTTCAAGCAGCTAGTG 1194
Db 302 LeuGlyAsnAspIleAlaLeuMetLysLeuAlaGlyProLeuThrPheAsnGluMetIle 321
QY 1195 AAACCAAGTGTCTGCCCCAACCCAGGCATGATCTGCAGCCAGACAGCTCTCTGCTGAT 1254
Db 322 GlnProValCysLeuProAsnSerGluGluAsnPheProAspGlyLysValCysTrpThr 341
QY 1255 TCCGGTGGGGGCCACCGAGGAGAGGAGAGGAGAGCTCAGAGTCTGACGCTGCGCAG 1314
Db 342 SerGlyTrpGlyAlaThrGluAspGlyGlyAspAlaSerProValLeuAsnHisAlaAla 361
QY 1315 GTGCTTCTCATTCAGACACAG-----AGATGCAACAGCAGATATGCTATGACAACTG 1368
Db 362 ValProLeuIleSerAsnLysAspLeuGlnProGlnGlyArgValArgTrpHisHisLeu 381
QY 1369 ATCACACCGCCATGATCTGTGCGGCTTCTCCAGGGG---AACGTGATTTCTGCGCAG 1425
Db 382 -----ProLeuHisAlaLeuArgGlyLeuProAspGlyTrpArgTrpAsnSerCysGln 399
QY 1426 GGTGACAGTGGAGGCTCTGCTGCTTCAAGAACAAATATCTGCTGCTGATAGGGAT 1485
Db 400 GlyAspSerGlyGlyProLeuValCysGlnGluArgLeuTrpLysLeuValGlyAla 419
QY 1486 ACAAGTGGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545
Db 420 ThrSerPheGlyIleGlyCysAlaAspValAsnLysProGlyValTyrThrArgValThr 439
QY 1546 GTATTACAGCTGGATTTATTCACAAATGAGGGCAGAC 1584
Db 440 SerPheLeuAspTrpIleHisGluGlnMetGluArgAsp 452

RESULT 10

US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, A
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.0

; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Alignment Scores:

Pred. No.: 2,98e-52 Length: 423
Score: 684.00 Matches: 165
Percent Similarity: 50.22% Conservative: 64
Best Local Similarity: 36.18% Mismatches: 167
Query Match: 21.52% Indels: 60
DB: 4 Gaps: 16

US-09-323-597C-1 (1-1738) x US-09-656-002-2 (1-423)

QY 292 TCCAAACCCCTGCTGTCAGCCAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAG 351
DB 2 SerAsnPro-----CysAlaAsnProValSerPro---TrpArgProSerGluSerVal 18
QY 352 ACTNAGAAAGCAGCTGTCATCACCCTGACCCCTGGGACCTTCCTCGTGGAGCTGGCTG 411
DB 19 GlyIleProIleIleAlaLeuLeuSerIleIleIleValValVal 38
QY 412 GCCCTGCTTACTCTGCAAGTTCATGGCAGCAAGTCTCCAACTCTGGGATAGACTGC 471
DB 39 IleIysValIleLeuAspLysTyr-----PheLeuCys 50
QY 472 GACTCCTCAGTACTCATCAACCCCTTAACCTGTTGATGCGGTGTCACACTGCCCC 531
DB 51 GlyIleProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCysPro 70
QY 532 GCGCGGAGCAGAGATCGGTGTGCTGCTCTAC-----GGACCA----- 573
DB 71 LeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAlaVal 90
QY 574 -----AAGTTCATCTCAGGTGTACTCATCTCAGAGAGTCTGTCAC 618
DB 91 ArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrpPhe 110
QY 619 CCTGTGTCGAAGCAGTGAAGCAGACTACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 678
DB 111 SerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgLysMetGly 130
QY 679 TAT-----AAGTAATATTTTACTCTAGCCAA-----GGATA 711
DB 131 TyrSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLeuAspVal 150
QY 712 GTGATGACAGCGGATCCAGCTTTTATGAATGAACAGTGAACAGTCCCGGCAATGTGAT 771
DB 151 ValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGly----- 167
QY 772 ATCTATAAAACGTACACAGTATGCTGCTTCTTCAAAAGCAGTGGTTTCTTTACGC 831
DB 168 -----ProCysLeuSerGlySerLeuValSerLeuHis 178
QY 832 TGTATAGCTCGGGGTCAACTGAATCAAGCCCGCAGACAGATGTGGCGGCGAG 891
DB 179 CysLeuAlaCysGlyLysSerLeu-----LysThrProArgValValGlyGlyGlu 195
QY 892 AGCGCGTCCCGGGGCTGCGCTGCGAGTGCAGTGCAGTCCAGAGCTCCAGCTG 951
DB 196 GluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGluHisVal 215
QY 952 TCGGAGGCTCCATCATCACCCTCGAGTGTGATGACAGCGGCGGCTGCTGCTGAGAAA 1011
DB 216 CysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLys 235
QY 1012 CCTCTTAACATCATGATGAGCGGCATTTGCGGGGATTTGAGA---CAATCTTTC 1068
DB 236 His---ThrAspValPheAsnTrpLysValArgAlaGlySerAspLysLeuGlySerPhe 254
QY 1069 ATGTTCTATGAGCGGATACCAAGTAGAAAAAGTGATT-----TCTCATCCCAAT 1119

DB 255 -----ProSerLeuAlaValAlaLysIleIleIleIleIleGluPheAsnProMet 270
QY 1120 TATGACTCCAGACCAAGAACATGACATTGGCGCTGATCAAGCTGCAGAGCCTCTGACT 1179
DB 271 Tyr-----ProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThr 287
QY 1180 TTCAAGACCTAGTGAACACAGTGTCTGCTGCCCAACCCAGGCAATGCTGCAGCCAGAA 1239
DB 288 PheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAla 307
QY 1240 CAGCTCTGCTGATTTCCGGTGGGGGCCACCCAGAGAGAAA---GGGAGAGCTCAGAA 1296
DB 308 ThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAsp 327
QY 1297 GTGCTGAAGCTGCCAAGGTCTTCTCATTTGACACACAGATGACACACATATGTC 1356
DB 328 IleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAla 347
QY 1357 TATGACAACTGATCACACAGCCATGATCTGTCGGGTCTCTCGAGGGGAAACGTCGAT 1416
DB 348 TyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAsp 367
QY 1417 TCTTGCAGGGTGACAGTGGAGGCTCTGCTCCTCCTCAATTCGAAAGAACATATCTCGTGGCTG 1476
DB 368 ThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln---TrpHisVal 386
QY 1477 ATAGGGGATACAGCTGGGGTCTGCTGCTGCCAAAGCTTACAGACAGGAGTGTACGGG 1536
DB 387 ValGlyLysValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThr 406
QY 1537 AATGTGATGGTATTCAGGACTGATTTATTCGACAAATGAGGCGAGAC 1584
DB 407 LysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGlu 422

RESULT 11

US-09-851-588-6
; Sequence 6, Application US/09851588
; Patent No. 6682890

GENERAL INFORMATION:

; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AI
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; PRIORITY FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-6

Alignment Scores:
Pred. No.: 1,35e-51 Length: 406
Score: 676.50 Matches: 150
Percent Similarity: 53.91% Conservative: 57
Best Local Similarity: 39.06% Mismatches: 128
Query Match: 21.28% Indels: 49
DB: 4 Gaps: 13

US-09-323-597C-1 (1-1738) x US-09-851-588-6 (1-406)

QY 508 TGTGATGGCTGTACACATCTCCCGGGGAGGACGAGATCGGTGTGTTCGCTCTAC 567
DB 46 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 65

QY 568 -----GGACCA-----AACTCATCTCTCAGGTGAC 594
Db |||||
66 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 85
QY 595 TCATCTCAGAGAGAGTCTCGCACCTGTGTGCCAAGACGACGTGAAGACGAGAACTACGGG 654
Db |||||
86 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 105
QY 655 CGGGCGGCTCGAGGACATCGGTAT-----AAGATAATTTTACTCTAGCCAA--- 705
Db |||||
106 GluThrAlaCysArgGlnMetGlyTySerSerLysProThrPheArgAlaValGluLeu 125
QY 706 -----GGAAATGATGATGACACCGGATCCACCGCTTTATGAACCTG 747
Db |||||
126 GlyProAspGlnAspLeuValValGluLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 145
QY 748 AACCAAGTCGGCGAATGTCGATATCTATAAAACGTACACAGNATGCTGTCT 807
Db |||||
146 ArgAsnSerSerGly-----ProCysLeu 153
QY 808 TCAAAGCAGTGTCTTTTACGCTGTATAGCTGCGGGGTCAACTTGAACCTCAAGCGC 867
Db |||||
154 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 170
QY 868 CAGACGAGTGTGGCGGAGAGCGGCTCCCGGGGCTGCGCTGCAGGTGACG 927
Db |||||
171 ThrProArgValValGlyGlyGlnGluAlaSerValAspSerTrpProTrpGlnValSer 190
QY 928 CTGACGCTCCAGACGTCACGCTGTGCGAGGCTCCATCATCACCCCGAGTGGATCGTG 987
Db |||||
191 IleGlnTyAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 210
QY 988 ACAGCGCCGCTGCTGGTGAACCTCTTAACTTCCATGCGCATGACCGCATTTGCG 1047
Db |||||
211 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 229
QY 1048 GGGATTTTGAGA---CAATCTTTCATGTCTATGAGCGCGATACCAAGTAGAGAAAGCTG 1104
Db |||||
230 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 245
QY 1105 ATT-----TCTCATCCAAATATGACTCCAGACCAAGACCAATGACATTCGCGTG 1155
Db |||||
246 IleIleIleGluPheAsnProMetTy-----ProLysAspAsnAspIleAlaLeu 262
QY 1156 ATGAGCTGCAGACCTCTGACTTTCAACGACCTAGTGAACCGAGTGTCTGCCCAAC 1215
Db |||||
283 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 282
QY 1216 CCAGCATGATCTGCACGACAGCTCTGCTGTGATTTCCGGTGGGGGCCACCGAG 1275
Db |||||
283 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 302
QY 1276 GAGAAA---GGAGACCTCAGAGTGTGACCGCTGCAAGGTGCTTCTTCATTGAGACA 1332
Db |||||
303 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 322
QY 1333 CAGAGATCAACAGCAGATATGCTTATCAACACCTGATCACACCGACCGATGCTGCC 1392
Db |||||
323 ThrArgCysAsnAlaAspAlaTyGlnGlyGluValThrGluLysMetMetCysAla 342
QY 1393 GGCTTCTCGAGGGAAGCTGATCTTCGAGGTGACAGTGGAGGCGCTCTGCTCACT 1452
Db |||||
343 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTy 362
QY 1453 TCGAAGACAAATATCTGCTGCTGATAGGGATACAACTGGGGTTCCTGGCTGCGCCAAA 1512
Db |||||
363 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyGlyCysGlyGly 381
QY 1513 GCTTACAGACGAGGAGTACGGGAATGTGATGATTTCAGGACTGGATTTATCGACAA 1572
Db |||||
382 ProSerThrProGlyValTyThrLysValSerAlaTyLeuAsnTrpIleTyAsnVal 401

QY 1573 ATGAGGCGACAC 1584
Db 402 TrpLysAlaGlu 405
RESULT 12
US-09-008-271A-6
Sequence 6, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLN0T13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6
Alignment Scores:
Pred. No.: 1,39e-51 Length: 435
Score: 676.50 Matches: 150
Percent Similarity: 53.91% Conservative: 57
Best Local Similarity: 39.06% Mismatches: 128
Query Match: 21.28% Indels: 49
DB: 3 Gaps: 13
US-09-323-597C-1 (1-1738) x US-09-008-271A-6 (1-435)
QY 508 TGTGATGGGTGTGCACACTGCCCGCGGAGGACGAGAAATCGTGTGTCCTCCTAC 567
Db |||||
75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluLysCysValLysSerPhe 94
QY 568 -----GGACCA-----AACTCATCTCTCAGGTGAC 594
Db |||||
95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114

```
QY 595 TCATCTCAGAGGAGTCTCTGGACCCCTGTGTGTCAGACGCACTGGACAGAACTACGGG 654
Db 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnThrGluAlaLeuAla 134
QY 655 CGGCGCGCTCGCAGGACATGGGCTAT-----AAGAATAATTTTACTCTACCCAA--- 705
Db 135 GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle 154
QY 706 -----GGAAATAGTGATGACAGCGGATGCCACCGCTTATTAAGAACTG 747
Db 155 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 174
QY 748 AACCAAGTCCGCGCAATGTCGATATCTATAAAAACTGTACCACATGATGCTCTTCT 807
Db 175 ArgAsnSerSerGly-----ProCysLeu 182
QY 808 TCAAAAGCACTGTTTTCACCTGTATAGCTCGGGGTCACTTGAATTTGAACCTG 867
Db 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyGluSerLeu-----Lys 199
QY 868 CAGAGCAGGATCTGGCGCGCAGAGCGCGCTCCCGGGCGCTGGCGCTGGCAGGTACG 927
Db 200 ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer 219
QY 928 CTGCACGTCAGAACCTCCACGCTGTCGGAGGCTCCATCATCACCCCGAGTGATG 987
Db 220 IleGlnTyrAspLysGlnHisValCysGlyGlySerLysLeuAspProHisTrpValLeu 239
QY 988 ACAGCGCCGACCTGGCGTGGAAAACTCTTAACAATCCATGCGATGGAGCGCATTTGG 1047
Db 240 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 258
QY 1048 GGGATTTTGA---CAATCTTTTCATGTTCTATGAGCGCGATACCAAGTAGAAAAAGT 1104
Db 259 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 274
QY 1105 ATT-----TCTCATCAAAATTATGACTCCAGACCCAGAACCAATGACATTCGCGT 1155
Db 275 IleIleIleGluPheAsnProMetTyr-----ProlAspAsnAspIleAlaLeu 291
QY 1156 ATGACGCTCAGAGGCTCTGACTTTCAACGACTAGTGAACCACTGTCTGCCCAAC 1215
Db 292 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 311
QY 1216 CCAGCATGATGTCGACGACAGAACAGCTCTGCTGATTTCCGGGTGGGGGCCACCGAG 1275
Db 312 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 331
QY 1276 GAGAAA---GGGAGACCTCAGAGTGTGCAACGCTGCCAAGTGTCTTCATTTGAGACA 1332
Db 332 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 351
QY 1333 CAGAGATGCAACAGCAGATATGCTATGACAACCTGATCAGACCCATGATCTGTGCT 1392
Db 352 ThrArgCysAsnAlaAspAlaThrGlnGlyGluValThrGluLysMetMetCysAla 371
QY 1393 GCTTCCTCGAGGGGAACGTGATCTTCACAGGTGACAGTGGAGGCTCTGTGTCAT 1452
Db 372 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 391
QY 1453 TCGAAGAACATATCTGGTGGCTGATAGGGGATACAGCTGGGGTCTCGCTGTCGCAA 1512
Db 392 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGly 410
QY 1513 GCTTACAGACCGAGTGTACGGGAATGTGATGTTATTCACGAGCTGGATTTATCGACAA 1572
Db 411 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 430
QY 1573 ATGAGGGGACAC 1584
Db 431 TrpLysAlaGlu 434
```

RESULT 13

```
US-09-851-588-8
; Sequence 8, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-68823-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-851-588-8

Alignment Scores:
Pred. No.: 1,39e-51 Length: 437
Score: 676.50 Matches: 150
Percent Similarity: 53.91% Conservative: 57
Best Local Similarity: 39.06% Mismatches: 128
Query Match: 21.28% Indels: 49
DB: Gaps: 13

US-09-323-597C-1 (1-1738) x US-09-851-588-8 (1-437)
QY 508 TGTGATGGGTCTCACACTGCCCCGGCGGGAGGACGAGATCGGTGTGTCGCTCTAC 567
Db 77 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 96
QY 568 -----GACACA-----AACTTCATCTCTCAGGTGTAC 594
Db 97 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 116
QY 595 TCATCTCAGAGAACTCTCTGCACCTGTGTGCCAAGACGACTGGAACGAGAACTACGGG 654
Db 117 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnThrGluAlaLeuAla 136
QY 655 CGGCGCGCTCGCAGGACATGGCTAT-----AAGAATAATTTTACTCTAGCCAA--- 705
Db 137 GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle 156
QY 706 -----GGAAATAGTGATGACAGCGGATCCACCGCTTTATGAACCTG 747
Db 157 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 176
QY 748 AACCAAGTCCGCGCAATGTCGATATCTATAAAAACTGTACCACATGATGCTCTTCT 807
Db 177 ArgAsnSerSerGly-----ProCysLeu 184
QY 808 TCAAAAGCACTGTTTTCATCGCTGTATAGCTCGCGGGTCACTTGAACCTCAAGCCGC 867
Db 185 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 201
QY 868 CAGAGCAGGATTTGGCGCGGAGAGCGCGCTCCCGGGCGCTGCGCTCGCAGGTCTAGC 927
Db 202 ThrProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSer 221
QY 928 CTGCACGTCAGAACGCTCCACGCTGTGCGAGGCTCCATCATCACCCCGAGTGATGCTG 987
Db 222 IleGlnTyrAspLysGlnHisValCysGlyGlySerLysLeuAspProHisTrpValLeu 241
QY 988 ACAGCGCCGACCTGCTGGAAAAAACCCTTTAAACAATCCATGCGATGGAGCGCATTTGC 1047
Db 242 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 260
```

```
QY 1048 GGGATTTTGA---CAATCTTTCATGTTCTATGAGCGCCGATACCAAGTAGAAGAAAGT 1104
D 1111
Db 261 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 276
QY 1105 ATT-----TCTCATCCAAATTTATGACTCCAGACCAAGAACAAATGACATTCGCGT 1155
D 1111
Db 277 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 293
QY 1156 ATGAAGCTCGAAGACCTCTGACTTTCAACGACCTAGTGAACCAAGTGTGTCTCCCAAC 1215
D 1111
Db 294 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 313
QY 1216 CCAGGCATGATGCTGCACCGACAGACCTCTGCTGGATTCGCGTGGGGGGCCACCGAG 1275
D 1111
Db 314 PheAspGluGluLeuThrProAlaThrProLeuThrPheIleIleGlyTyrGlyPheThrLys 333
QY 1276 GAGAAA---GGGAAGACCTCAGAAGTGTCTGAACGCTGCCAAGGTCTTCTCATTTAGACA 1332
D 1111
Db 334 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 353
QY 1333 CAGGATGTCACACGAGATATGTCTATGACAACTGATCACCAGCCATGATCTGTGCC 1392
D 1111
Db 354 ThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAla 373
QY 1393 GGCTTCTCTCGCAGGGGAAGCTCGATTCTTGCCAGGCTGACAGTGGAGGCGCTCTGTCAC 1452
D 1111
Db 374 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 393
QY 1453 TCGAAGAACATATCTGTGCTCATAGGATACAGCTGGGTTCTGGCTGTGCCAAA 1512
D 1111
Db 394 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGly 412
QY 1513 GCTTACAGACGAGGATGTACGGAAATGTGTGATTCTCCGAGTGTGATTATTCGACAA 1572
D 1111
Db 413 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 432
QY 1573 ATGAGGGCGAGAC 1584
D 1111
Db 433 TrpLysAlaGlu 436
```

RESULT 14

```
US-09-323-597c-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Alignment Scores:
Pred. No.: 5,01e-50 Length: 798
Score: 660.00 Matches: 150
Percent Similarity: 48.98% Conservative: 67
Best Local Similarity: 33.86% Mismatches: 180
Query Match: 20.76% Indels: 46
DB: 12 Gaps: 12

US-09-323-597c-1 (1-1738) x US-08-200-900A-2 (1-798)
QY 304 GTCTGCACGACGAGCCAAATCCCATCCGACAGTGTGCACCTCAAGACTAAGAAGCA 363
D 1111
Db 374 ValTyrThrGlyProGlyProValAsnAspValPheSerThrThrAsnArgMetThrVal 393
QY 364 CTGTGCATCACCTGTACCCCTG-----GGGACCTTCTCTCGTGGGA 402
D 1111
Db 394 LeuPheIleThrAspAsnMetLeuAlaLysGlnGlyPheLysAlaAsnPheThrThrGly 413
QY 403 GCTGCGCTGCGCGCTGCTCTCTGAAAGTTCATGGCGCAGCAAGTCTCCAACTCTGGG 462
D 1111
Db 414 TyrGlyLeuGly-----IleProGluProCysLysGluAspAsn 426
QY 463 ATAGAGTGGACCTCTCAGTACTCATCAACCCCTCTAACTGTGTGTGATGCGGTGCA 522
D 1111
Db 427 PheGlnCys---LysAspGlyGluCysIleProLeuValAsnLeuCysAspGlyPhePro 445
QY 523 CACTGCCCCGGCGGAGGAGCAATCGTGTGTGTCTCTACGGA-----570
D 1111
Db 446 HisCysLysAspGlySerAspGluAlaHisCysValArgLeuPheAsnGlyThrThrAsp 455
QY 571 -----CCAAACTTTCATCTCTCAGTGTACTCATCTCAGAGGAGTCTCTGGCAC 618
D 1111
Db 466 SerSerGlyLeuValGlnPheArgIleGln-----SerIleTyrHis 479
QY 619 CTTGTGTGCGACAGCACTGCGACAGAACTACGGCGCGCGCGCTCGAGGACATGGCC 678
D 1111
Db 480 ValAlaCysAlaGluAsnTrpThrThrGlnIleSerAspValCysGlnLeuLeuGly 499
QY 679 TATAAGATAATTTTACTCTAGCCAGGAATAGTCGATCAGACGCGATCCACGAGCTTT 738
D 1111
Db 500 LeuGlyThrGly-----AsnSerSerValProThrPheSerThrThrGlyGlyProTyr 517
QY 739 ATGAACCTGACACAAAGTCCCGCAATGTCGATATCTATAAAACCTGTACACAGTGTAT 798
D 1111
Db 518 ValAsnLeuAsnThrAlaProAsnGlySerLeuIle-----LeuThrProSerGln 534
QY 799 GCCTGTCTTCTCAAAAGCAGTGGTTCTTTACGCTGT-----ATAGCTCGCGGGTCAAC 852
D 1111
Db 535 GlnCysLeuGluAspSerLeuIleLeuGlnCysAsnTyrLysSerCysGlyLysLys 554
QY 853 TTGAACCTCAAGCCGCCAGAGC---AGGATTGTGGCGCGGAGAGCGCGTCTCCGGGGGCC 909
D 1111
Db 555 LeuValThrGlnGluValSerProLysIleValGlyGlySerAspSerArgGluGlyAla 574
QY 910 TGGCCCTGCGCAGTCTGACGCTCCAGACGTCACAGTCCACGCTGTGCGGAGGCTCCATCATC 969
D 1111
Db 575 TrpProTrpValValAlaLeuTyrPheAspAspGlnGlnValCysGlyAlaSerLeuVal 594
QY 970 ACCCCGAGTGTGATCGTACAGCCGCCCTCTGCTGTGGAAGAAACCTCTTAACTCATGTG 1029
D 1111
Db 595 SerArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnMetGluProSer 614
QY 1030 CATTGACGCGCATTTCCGGGATTTTGAGACAACTCTTTCATGTTCTATGAGCCGGATAC 1089
D 1111
Db 615 LysTrpLysAlaValLeuGlyLeuHisMetAlaSerAsnLeu-----ThrSerPro 631
```

Qy 1090 CAAGTAGAA-----AAAGTGAATTTCTCATCAAAATATGACTCCAGACC 1134
Db 632 GlnileGluThrArgLeuileAspGlnileValileAsnProHisTyrAsnLysArgArg 651
Qy 1135 AAGAACATGATGCTGCTGATGAGCTGAGAGCTCTGACTTTCAAGCAGCTAGTG 1194
Db 652 LysAsnAsnAspIleAlaMeCMeHisLeuGluMetLysValAsnTyrThrAspTyrIle 671
Qy 1195 AAACAGTGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTGGATT 1254
Db 672 GlnProileCysLeuProGluGluAsnGlnValPheProGlyArgIleCysSerile 691
Qy 1255 TCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGAGCTGCCAAG 1314
Db 692 AlaGlyTrpGlyAlaLeuileTyGlnGlySerThrAlaAspValLeuGlnGluAlaAsp 711
Qy 1315 GTGCTTCTCATGAGACAGAGATGCAACAGCAGATATGCTATGCAACCTGATCACA 1374
Db 712 ValProLeuLeuSerAsnGluLysCysGlnGlnGlnMetProGluTyrAsn---IleThr 730
Qy 1375 CCAGCCATGATGCTGCGGCTTCTCGAGGGGAGAGCTGATCTTCCAGGCTGACAGT 1434
Db 731 GluAsnMetValCysAlaGlyTyGlnAlaGlyValAspSerCysGlnGlyAspSer 750
Qy 1435 GGAGGCTCTGGTCACTTCAAGCAATATCTGGTGGCTGATGAGGGATACAGCTGG 1494
Db 751 GlyGlyProLeuMetCysGlnGluAsnAsnArgTrpLeuLeuAlaGlyValThrSerPhe 770
Qy 1495 GGTTCGTGTGCTGCAAGCTTACAGACAGAGTGTACGGGAGATGATGATTCACG 1554
Db 771 GlyTyrGlnCysAlaLeuProAsnArgProGlyValTyrAlaArgValProArgPheThr 790
Qy 1555 GACTGGATT 1563
Db 791 GlnTrpIle 793

RESULT 15

PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

Alignment Scores:
Pred. No.: 5.01e-50 Length: 798
Score: 660.00 Matches: 150
Percent Similarity: 48.98% Conservative: 67
Best Local Similarity: 33.86% Mismatches: 180
Query Match: 20.76% Indels: 46
DB: 5 Gaps: 12

US-09-323-597C-1 (1-1738) x PCT-US94-00616-2 (1-798)

Qy 304 GTCTGCACGGCCCAATCCCATCGGGACAGTGTGCACCTCAAGACTAAGAAAGCA 363

Db 374 ValTyrThrGlyProGlyProValAsnAspValPheSerThrThrAsnArgMetThrVal 393
Qy 364 CTGTGATCACCCTTGACCTG-----GGGACCTTCTCTCGTGGGA 402
Db 394 LeuPheIleThrAspAsnMetLeuAlaLysGlnGlyPheLysAlaAsnPheThrThrGly 413
Qy 403 GCTGCGCTGGCCCTGCTCTGAGAGTTCTATGGCAGCAAGTGTCTCAACTCTGGG 462
Db 414 TyrGlyLeuGly-----IleProGluProCysLysGluAspAsn 426
Qy 463 ATAGAGTGGACCTCTCAGTACCTCAGTACCTCAACCCCTCTAACTGCTGTGTATGCGGTGCA 522
Db 427 PheGlnCys---LysAspGlyGluCysIleProLeuValAsnLeuCysAspGlyPhePro 445
Qy 523 CACTGCCCCGGGGAGGAGACGAGATCGGTGTGTGCTCTACCGA 570
Db 446 HisCysLysAspGlySerAspGluAlaHisCysValArgLeuPheAsnGlyThrThrAsp 465
Qy 571 -----CCAAACTTCTCTCAGGTGTACTCTCAGAGGAAGTCTTGCAC 618
Db 466 SerSerGlyLeuValGlnPheArgIleGln-----SerIleTrpHis 479
Qy 619 CTGTGTGCAAGACGACTGGAACGAGAACTACGGCGCGCGCTCTCAGGGACATGGGC 678
Db 480 ValAlaCysAlaGluAsnTrpThrThrGlnIleSerAspAspValCysGlnLeuLeuGly 499
Qy 679 TATAAGNATAATTTTACTCTAGCCAAAGNATAGTGTGATGACAGACCGGATCCACGCTTT 738
Db 500 LeuGlyThrGly-----AsnSerSerValProThrPheSerThrThrGlyGlyProTyr 517
Qy 739 ATGAACTGCAACAAAGTCCGGCAATGTCTGATATCTATAAAACCTGTACACAGTGAT 798
Db 518 ValAsnLeuAsnThrAlaProAsnGlySerLeuIle-----LeuThrProSerGln 534
Qy 799 GCCTGTCTTCAAAACAGTGTCTTTCAGCTGT-----ATAGCTTCCGGGTCAAC 852
Db 535 GlnCysLeuGluAspSerLeuIleLeuLeuGlnCysAsnTyrLysSerCysGlyLysLys 554
Qy 853 TTGAACCTCAAGCGCCAGAGC---AGGATTTGGCGCGGAGAGCGCGCTCCCGGGGCC 909
Db 555 LeuValThrGlnLeuValSerProLysIleValGlySerAspSerArgGluGlyAla 574
Qy 910 TGGCCCTGGCAGGTCTGACAGCTGCAAGCTCCAGAGCTCCAGTGTGGGAGCTCCATC 969
Db 575 TrpProTrpValValAlaLeuTyrPheAspAspGlnGlnValCysGlyAlaSerLeuVal 594
Qy 970 ACCCGCGATGATCTGACAGCGCCCTGCTGTGAGAAACCTCTTAAACATCCATCG 1029
Db 595 SerArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnMetGluProSer 614
Qy 1030 CATTTGACCGGCATTTGCGGGGATTTTGAGACAATCTTTCATGTCTTATGAGCGCGGATC 1089
Db 615 LysTrpLysAlaValLeuGlyLeuHisMetAlaSerAsnLeu-----ThrSerPro 631
Qy 1090 CAAGTAGAA-----AAAGTGAATTTCTCATCAAAATATGACTCCAGACC 1134
Db 632 GlnileGluThrArgLeuileAspGlnileValileAsnProHisTyrAsnLysArgArg 651
Qy 1135 AAGAACATGATGCTGCTGATGAGCTGAGAGCTCTGACTTTTCAAGCAGCTAGTG 1194
Db 652 LysAsnAsnAspIleAlaMetHisLeuGluMetLysValAsnTyrThrAspTyrIle 671
Qy 1195 AAACAGTGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTGGATT 1254
Db 672 GlnProileCysLeuProGluGluAsnGlnValPheProGlyArgIleCysSerile 691
Qy 1255 TCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGAGCTGCCAAG 1314
Db 692 AlaGlyTrpGlyAlaLeuileTyGlnGlySerThrAlaAspValLeuGlnGluAlaAsp 711
Qy 1315 GTGCTTCTCATGAGACAGAGATGCAACAGCAGATATGCTATGCAACCTGATCACA 1374

Tue Jul 6 15:26:52 2004

```

Db      712 ValProLeuLeuSerAsnGluLysCysGlnGlnGlnMetProGluTyrAsn---IleThr 730
QY      1375 CCAGCCATGATCTGTGCGGCTTCCTGTCAGGGGAACGTCGATTCTTGCAGGTCACAGT 1434
Db      731 GluAsnMetValCysAlaGlyTyrGluAlaGlyValAspSerCysGlnGlyAspSer 750
QY      1435 GGAGGGCTCTGGTCACCTTGGAAAGCAATATCTGGCTGCTGATAGGGGATACAGCTGG 1494
Db      751 GlyGlyProLeuMetCysGlnGluAsnAsnArgTrpLeuLeuAlaGlyValThrSerPhe 770
QY      1495 GGTTCCTGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGTTATTACG 1554
Db      771 GlyTyrGlnCysAlaLeuProAsnArgProGlyValTyrAlaArgValProArgPheThr 790
QY      1555 GACTGGATT 1563
Db      791 GluTrpIle 793

```

Search completed: July 6, 2004, 15:09:29
 Job time : 42 secs